



Farmfit
Intelligence



HOW TO BEST USE PRIMARY FARM-LEVEL DATA FOR IMPACTFUL SMALLHOLDER ENGAGEMENT MODELS

Lessons learned from IDH Farmfit

OCTOBER 2021

1.0

THE IMPORTANCE OF FARM-LEVEL DATA FOR THE DEVELOPMENT OF WELL-FUNCTIONING SMALLHOLDER AGRICULTURE MARKETS

- 8 **1.1**
How farm-level data can contribute to improving smallholder farmer livelihoods
- 9 **1.2**
Data challenges in smallholder agriculture

5.0

CONCLUDING REMARKS

- 40 **5.1**
Key report takeaways
- 41 **5.2**
Next steps in the road to data-driven impact at farm level

2.0

IDH FARMFIT'S MISSION TO IMPROVE FARMER LIVELIHOODS AND THE ROLE OF DATA

- 11 **2.1**
Improving service delivery models for SHFs
- 17 **2.2**
IDH Farmfit's data-driven methodology to improve service delivery models

3.0

IDH FARMFIT'S KEY REASONS FOR PRIMARY FARM-LEVEL DATA COLLECTION

- 23 **3.1**
Primary data strengthens SDM analyses and the design of TA projects
- 25 **3.2**
Primary data generates comparable aggregate data for learning across SDMs
- 27 **3.3**
Primary data is key to evaluate the impact of interventions on farmer livelihoods

4.0

IDH FARMFIT'S METHODOLOGY, CHALLENGES, AND BEST PRACTICES TO PRIMARY DATA COLLECTION

- 30 **4.1**
Scope of the data collection
- 31 **4.2**
Standardization and modularity in survey
- 33 **4.3**
Adaptation of the survey to the SDM and local context
- 35 **4.4**
Data collection process



EXECUTIVE SUMMARY

THE IMPORTANCE OF HIGH-QUALITY FARM-LEVEL DATA TO IMPROVE SMALLHOLDER FARMER LIVELIHOODS

While smallholder farmers (SHFs) are key players in agricultural value chains, they often lack access to affordable and high-quality goods and services (e.g., inputs, credit, markets, and information) that would enable them to improve their farms and incomes. Key barriers, such as risks associated with smallholder farming, keep the private sector from fully investing in service provision to SHFs. This problem is compounded by a limited understanding of what works best for the private sector and for SHFs, due in large part to the limited quantity, quality, and interoperability of farm-level data. In this report, we argue that accurate farm-level data is critical to creating financially viable and investable smallholder engagement models that improve farmer livelihoods.

In Section 1 of this report, we discuss the importance of reliable data and the challenges of collecting it from SHFs, which are widely documented in agricultural markets. Challenges include logistical issues of reaching research participants, especially in rural or remote areas; lack of available farm-level records; complexity and vast differences in measurement approaches; costs of collecting primary data;

and the underdeveloped regulatory environment governing the collection, use, and sharing of farm-level data, which can contribute to reluctance among farmers to share information.

IDH FARMFIT'S MISSION TO IMPROVE SERVICE DELIVERY MODELS TO SHFS AND THE KEY ROLE OF DATA

In Section 2, we illustrate how and why data is critical for IDH Farmfit to achieve its mission, which is to improve farmer livelihoods by contributing to the transformation of smallholder agriculture markets. Through analytics, technical assistance, and catalytic investment funding, we support and guide the private sector in making their service delivery to SHFs profitable and investable while also improving farmers livelihoods.

We have developed a data-driven methodology to systematically analyse service delivery models (SDMs). This methodology helps SDM operators (e.g., input providers, off-takers, financial service providers) understand the conditions under which goods and services can be provided to SHFs in an efficient and effective way, with a sustainable return on investment. This in-depth assessment is critical to inform the design of our technical assistance (TA), which aims to help SDMs become “investment ready” and scalable to maximize

positive impact on farmer livelihoods. By investment ready, we mean the degree to which SDMs are able to attract financing to fund their operations and scaling. The value that an SDM creates for farmers is key to the decision-making process of potential investors; thus, high-quality farm-level data is critical to helping SDMs become investment ready.

Given that SHFs are the primary customer of SDMs—and, in many cases, are also a supplier of agricultural products to SDM operators—a solid understanding of their socio-economic characteristics and agronomic behaviour is crucial if we are to design and implement sustainable, scalable, and investable SDMs that enhance farmers' income and resilience.

HOW FARM-LEVEL DATA COLLECTION ADDS VALUE TO OUR WORK

In Section 3, we make the case that primary data collection is often the best way of generating accurate, relevant, and comparable data about SHFs. We also discuss the three main ways it adds value to our work, namely by:

1. **Strengthening the analysis of SDMs and inform the design of TA projects** with an aim of making SDMs efficient, effective, commercially viable, and investable at scale;
2. **Generating comparable aggregate data for comparison and learning** across SDMs, allowing various stakeholders to make informed decisions and calculated risks on investments in smallholder farming; and
3. **Measuring and evaluating the impact of interventions** on farmer livelihoods, to help steer interventions to maximize positive impact.

IDH FARMFIT'S METHODOLOGICAL APPROACH TO PRIMARY DATA COLLECTION, CHALLENGES, AND BEST PRACTICES

In Section 4, we provide an overview of our methodological approach to data collection. We discuss how a well-developed research design, data collection instrument, and collection process, helps us tackle challenges associated with agricultural data (such as a lack of quality, comparable, and interoperable farm-level data).

Critical to the design of our data collection instrument is the adaptation of our survey to local nuances or measurement units. In this way, we can collect data on a certain crop or in a geographical region, while ensuring that these questions remain standardized and comparable across SDM analyses. Each time we conduct data collection, our Question Library is enriched with additional options for tailored survey questions and answers. This rigorous approach not only ensures measurement accuracy, but also facilitates survey design for future data collection efforts.

In terms of the data collection process, we share best practices to ensure timely data collection and to generate high-quality data. These include the training of local enumerators and the use of digital tools to administer the survey and perform data quality checks. Importantly, these tools also allow for remote data collection, such as during the COVID-19 pandemic. Another important dimension of our data collection process is our codes of conduct for data ownership and privacy. While we ensure informed consent from survey participants, anonymization of data, and transparency on the use of data, we acknowledge that data ownership, privacy, and feedback mechanisms are delicate issues which require continuous attention and effort.

CALL FOR ACTION: THE ROAD TO DATA-DRIVEN IMPACT ON SHF LIVELIHOODS IS BEST TRAVELLED TOGETHER

Finally, we conclude this report with a call for action. We are keen on engaging with other development organizations and knowledge partners to optimize our data-driven journey, share insights and lessons, and create alignment in the industry. For example, we are eager to learn how others deal with data ownership and feedback mechanisms to SHFs. We seek to contribute to shared principles and approaches to data collection and analysis—for example, operationalizing concepts into industry relevant indicators. In the longer run, we aim to convene other organizations in the sector to use increasingly harmonized language and methodologies, and hope that the data and insights generated in this manner will become increasingly interoperable. As such, the ease and value of benchmarking can be continuously strengthened and the impact on farmer livelihoods can be maximized.

IN THIS REPORT, IDH FARMFIT WILL:

- Highlight the importance of data for improving SHF livelihoods, as well as key challenges;
- Explain the value of primary farm-level data:
 - For IDH Farmfit’s engagements with the private sector—specifically, for high-quality analyses of SDMs and informed TA to make SDMs investment ready, scalable, and effective;
 - For generating aggregated insights to drive change in service markets for SHFs; and
 - For evaluating the impact of interventions on farmer livelihoods and adapting interventions accordingly.
- Provide transparency on IDH Farmfit’s methodological approach to primary data collection at farm level, including challenges and best practices, thereby offering guidance and tools to others seeking to collect such data; and
- Invite our peers to engage with IDH Farmfit to optimize our data-driven journey, to share insights and create alignment in the industry, and to reflect on the future role of primary data in smallholder agriculture.





1.0

THE IMPORTANCE OF FARM-LEVEL DATA FOR THE DEVELOPMENT OF WELL-FUNCTIONING SMALLHOLDER AGRICULTURE MARKETS

-
- 1.1 HOW FARM-LEVEL DATA CAN
CONTRIBUTE TO IMPROVING
SMALLHOLDER FARMER LIVELIHOODS
 - 1.2 DATA CHALLENGES IN SMALLHOLDER
AGRICULTURE

1.1

HOW FARM-LEVEL DATA CAN CONTRIBUTE TO IMPROVING SMALLHOLDER FARMER LIVELIHOODS

Most of the world's farms are smallholdings (typically farms less than two hectares in size)¹ which are predominantly cultivated by the world's poor. Smallholder farmers (SHFs) face numerous challenges—such as a lack of access to inputs (e.g., seeds, fertilizers), markets, finance, and information—all of which limit their ability to generate sustainable livelihoods and build resilience. Improving agricultural productivity and household incomes is central to reducing global poverty and food insecurity. This is recognized in the Sustainable Development Goals (SDGs)—particularly in SDG 2 with its focus on ending hunger, achieving food security, and promoting sustainable agriculture.

The collection and use of high-quality data is central to achieving these goals. Development organizations need data to tailor their interventions to SHFs, set measurable targets, and monitor progress. Yet detailed data on the impact of interventions focused on transforming smallholder agriculture markets is rarely available; what data is available is often incomplete, not easily comparable to other contexts, or of insufficient quality.²

High-quality data, especially at the farm-level, is also key to attract and de-risk private sector investments in smallholder agriculture markets. Significant barriers keep the private sector from fully investing in service provision to SHFs. These barriers include high costs and complexity of building effective business models, real and perceived risks associated with smallholder farming, and overall challenging economics of smallholder farming, which limit what farmers can pay for goods and services. This problem is compounded by a limited understanding of farmers' needs and of the impact of services on farmer livelihoods—due in large part to the poor quantity, quality, and interoperability of farm-level data. Farm-level data is thus critical to overcoming these challenges. It provides the private sector a better understanding of the needs of SHFs, the impact of interventions on farm-level profitability and resilience, how the risks and costs of service provision to SHFs can be reduced, and how they can adapt their SHF engagement models accordingly.

“Farm-level information is important for all actors in the agriculture supply chain—farmers, agricultural businesses, and their funders. Accurate information on farms and farm economics is essential to design support for farmers and improve its impact. This design process often happens in a vacuum, resulting in interventions that don't actually meet farmers' needs. Without accurate tracking of how farmers are experiencing the intervention and kinds of positive and negative impacts it is having in their lives, it's hard to ensure that the end product is having the desired effect.”

– Reflection from 60 Decibels*

* 60 Decibels is a customer centric impact measurement company. For more information see <https://60decibels.com/>

1. Lowder, S., Scoet, J., & Raney T. (2016). *The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide*, World development

2. FAO (2018). *Transforming food and agriculture to achieve the Sustainable Development Goals: 20 actions to implement the 2030 Agenda*. Accessed via <http://www.fao.org/3/I9900EN/i9900en.pdf>

1.2

DATA CHALLENGES IN SMALLHOLDER AGRICULTURE

Challenges to the collection of reliable data from SHFs are widely observed and documented in agriculture markets.³ First, SHFs can be large in number and geographically dispersed, creating logistical issues in reaching research participants, especially in rural or remote areas.⁴ Second, measuring accurate information on variables—such as production, yield, and farm size—is frequently complex and costly as farmers may not keep records or use standard measurement units. Depending on the geographical area where crops are grown, approaches to estimating production may differ, making comparison and aggregation of data difficult. Third, while a lack of quality data contributes to a lack of investment in service provision to SHFs, the inverse is also true. Even in cases where data is collected, it is often not in the form required; nor do many companies have the capacity to appropriately analyse the data. Therefore, companies often do not generate the kind of analytical insights required to adequately assess the achieved impact (or lack thereof) of their activities; best practices that can be replicated or learned from; or operational and strategic approaches that work better or worse in a particular context.

Due to these challenges, the value of primary data is often not visible enough to outweigh the cost of data collection. The digitalization of agriculture may help overcome some of the challenges related to cost and quality of data. Technology-driven data collection—for instance, using mobile or phone surveys, geospatial technologies for data collection, and farm management information systems (FMIS)—allows for data to

be collected with a higher level of efficiency, scale, and often quality. The adoption and benefits of FMIS in smallholder agriculture is discussed in IDH Farmfit's recent Insights Brief, *A practical guide for integrating data into farmers' decision making - lessons from Asia*.⁵

At the same time, the digitalization of agriculture has also led to emerging concerns around the relationship between primary data collection and trends related to data collection, management, and ownership. Farmers may be reluctant to share information because of data privacy and security concerns, or because they are not aware of the value it could offer them.⁶ Other issues include potential power imbalances, information asymmetries, and a lack of literacy among farmers (and often a complete absence of contractual agreements); these can leave farmers with feelings of disempowerment, and with little negotiating power or agency. In many contexts, the regulatory environment governing the collection, use, and sharing of data from farmers is undeveloped or not properly enforced.⁷ Thus, we need to protect the privacy and security of farmer data, address their other concerns, and ensure farmers also benefit from data collection, while also meeting the economic interest of stakeholders and leveraging data for the sector's growth and innovation.

3. ISF advisors (2020) *Digital data sharing in agriculture: Mercy Corps Agrifin case study*

4. Keita & Carfagna (2010). *Overview of methodological issues for research to improve agricultural statistics in developing countries*.

5. Accessible via <https://www.idhsustainabletrade.com/publication/a-practical-guide-for-integrating-data-into-farmers-decision-making-lesson-from-asia/>

6. Jouanjean, M., et al. (2020). *Issues around data governance in the digital transformation of agriculture: The farmers' perspective*, OECD Food, Agriculture and Fisheries Papers, No. 146, OECD Publishing, Paris

7. Zampati (2019). "Does data mean power for smallholder farmers?". Accessed via <https://blogs.worldbank.org/opendata/does-data-mean-power-smallholder-farmers>

2.0

IDH FARMFIT'S MISSION TO IMPROVE FARMER LIVELIHOODS AND THE ROLE OF DATA

2.1 IMPROVING SERVICE DELIVERY MODELS

**2.2 IDH FARMFIT'S DATA-DRIVEN
METHODOLOGY TO IMPROVE SERVICE
DELIVERY MODELS**

2.1

IMPROVING SERVICE DELIVERY MODELS FOR SHFS

IDH has invested in interventions to improve smallholder livelihoods since its inception. The IDH Farmfit program aims to make investment in smallholder farming more attractive and seeks to contribute to the transformation of smallholder agriculture markets by providing business support and de-risked finance models to the private sector, and evidence-based insights to a wider public.⁸ We believe that improvements in farmer livelihoods depend on markets that farmers can participate in on an equitable, participatory, and commercially viable way, while ensuring social and environmental outcomes. Our thinking on agricultural market transformation, the role of private sector development, and the strategies IDH Farmfit uses to support such transformation are the subject of a forthcoming Insights Brief, *Private Sector Development Strategies: Leveraging the Private Sector for Market Transformation in Smallholder Agriculture*.

In 2015, IDH realized that—despite global dedication of resources to technical assistance, grant funding, and blended finance for smallholder agriculture—not enough data-driven insights were available on what works for the private sector and the farmers they are engaged with.



Therefore, we developed a data-driven methodology to systematically analyse service delivery models (SDMs – see Box 1 for a definition and conceptualization of a typical SDM). This methodology is used by IDH Farmfit to provide an in-depth assessment of the business models of SDMs and the value proposition for SHFs, thereby creating a basis for informed decision making and improvement.

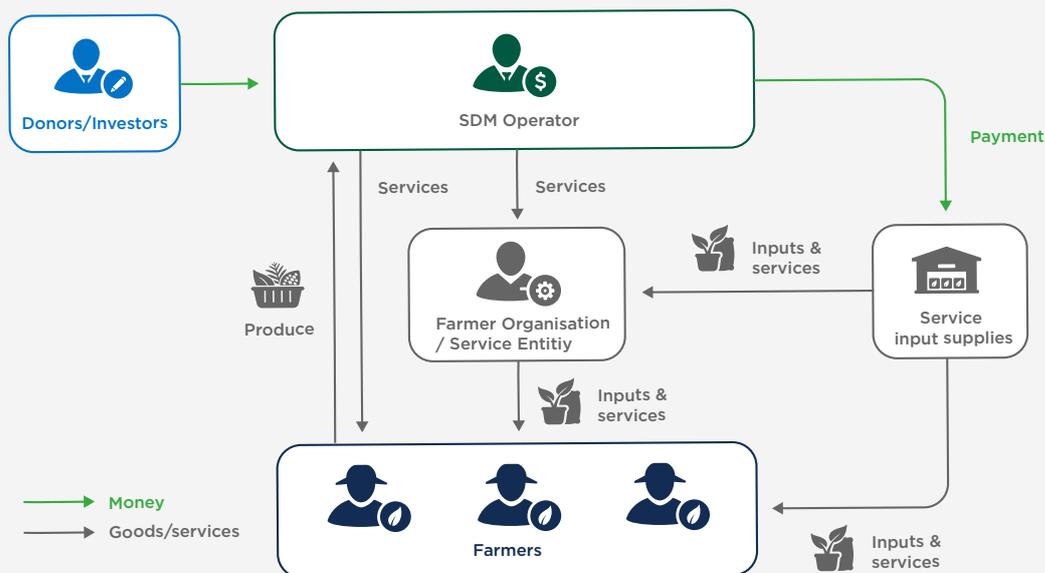
8. The IDH Farmfit program consists of three pillars: 1) business analytics and technical assistance to the private sector to improve smallholder engagement models (through Farmfit Business Support), 2) evidence-based insights on how to make smallholder value chains more efficient, effective, and impactful (through Farmfit Intelligence; and 3) de-risking and unlocking farmer financing and catalyzing commercial investments in working capital and asset finance to scale businesses (through the Farmfit Fund).

BOX 1 - SERVICE DELIVERY MODELS (SDMS)

SDMs are supply chain structures that provide services such as training, access to inputs, markets, and finance to farmers. SDMs can be operated by a range of organizations, including input providers, traders, processors, financial service providers, ag-tech and fin-tech providers, farmer organizations, NGOs, public extension schemes, or others; these are referred to as

SDM operators. IDH Farmfit examines the SDM from the perspective of the SDM operator, farmers, and other intermediaries and service providers involved in the SDM. We believe that SDMs that provide farmers with access to goods and services can play a key role in improving farmer livelihoods.

FIGURE 1: A BASIC SERVICE DELIVERY MODEL STRUCTURE



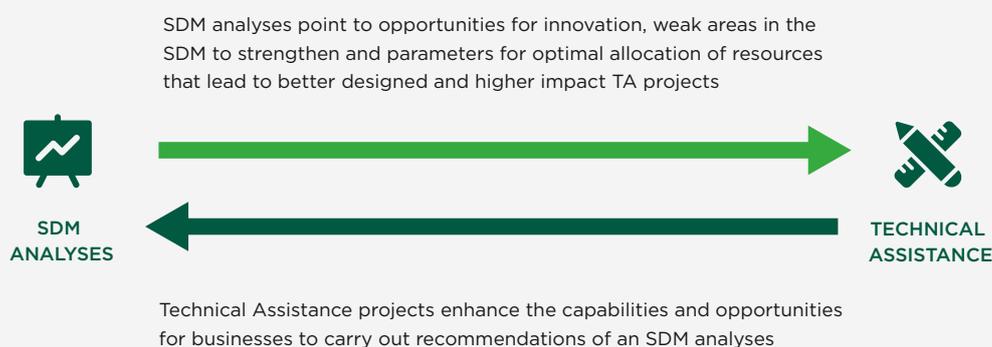
The SDMs that IDH Farmfit performs help SDM operators understand the conditions under which services (such as training, access to inputs, markets, finance, and information) can be provided to SHFs in an efficient and effective way, with a sustainable return on investment. This in-depth assessment lays the foundation for the design of technical assistance (TA) that IDH Farmfit provides in most of its private sector engagements. This technical assistance aims to strengthen the SDM so that it becomes “investment ready” and scalable, thus maximizing impact on farmer livelihoods (see Box 2 for an elaboration on IDH’s business support and the role of primary data).

BOX 2 - IDH FARMFIT PROVIDES BUSINESS SUPPORT TO SDM OPERATORS TO SCALE THEIR MODELS

IDH Farmfit’s business support to the private sector consists of SDM analyses, which are typically followed by the provision of technical assistance (TA) to SDM operators. SDM analyses provide critical insights for designing TA projects, but many businesses do not have the capabilities, resources or the risk appetite that would allow/incentivize them to carry out recommendations independently. Our TA interventions typically have a duration of two to three years during

which we guide and support SDM operators to optimize business management and operational efficiency, link them to markets and financiers, and provide co-funding and support for project setup and implementation. During the course of TA projects we monitor progress and gather primary data at both SDM and farmer levels to develop proof of concept and to test our innovations and gather learnings.⁹

FIGURE 2: IDH FARMFIT’S MUTUALLY REINFORCING INTERVENTIONS OF SDM ANALYSES AND TA

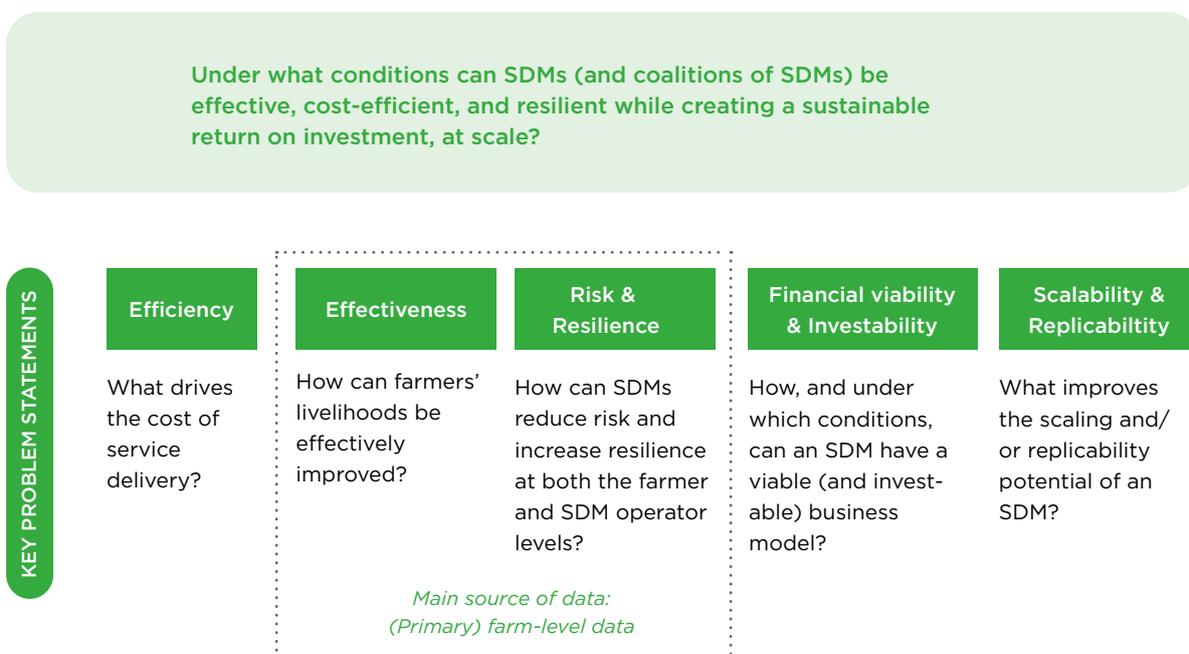


The collection of high quality data, including farm-level data, is the starting point of our private sector engagements. We consider this critical to developing impactful SDMs, and thus critical to our contribution to the sustainable transformation of smallholder agriculture markets.

9. Alongside our support to the private sector, the IDH Farmfit Fund invests, de-risks, and catalyses investments in smallholder-inclusive businesses and value chains. The Fund seeks to demonstrate that farmer financing is possible in a sustainable way by building a large and sound farmer-centric investment portfolio and increasing the availability of affordable, long-term financing to farmers. SDM analyses and TA can be provided pre-investment to optimize the investment readiness of the SDM operator, or post-investment to mitigate risks through TA interventions that enhance the likelihood of materializing a successful financial deal. (More information is available via www.idhsustainabletrade.com/farmfit-fund).

Figure 3 depicts our Learning Framework, which organizes and structures the areas in which we seek to generate, analyse, and disseminate insights in order to realize our mission. The overarching question we seek to answer is: *Under which conditions can SDMs (and coalitions of SDMs) be effective, cost-efficient, and resilient while creating sustainable return on investment, at scale?* To answer this question, we identify five underlying problem statements, each of which looks at a different component of well-functioning SDMs. For each of these problem statements, we have defined several hypotheses, which we test based on data points that we collect during SDM analyses and other private sector engagements.¹⁰

FIGURE 3: IDH'S FARMFIT LEARNING FRAMEWORK GUIDING THE COLLECTION OF DATA FOR OUR PRIVATE SECTOR ENGAGEMENTS



10. Examples of our hypotheses include: 1) SDMs where farmers are organized on contiguous/consolidated land are more efficient; 2) SDMs that provide value-adding services (storage, aggregation, transport) increase farmer incomes; 3) SDMs that incorporate sourcing are more financially sustainable than SDMs focusing solely on service provision; 4) SDMs that offer farmers a guaranteed market price reduce attrition rates by lowering the financial risks faced by farmers; and 5) SDMs operating in markets with an unsupportive enabling environment are less able to achieve scale. More information on the hypotheses underpinning each problem statement can be shared by IDH Farmfit upon request.

Although our focus is on developing and strengthening SDMs (which requires data at the SDM operator level), for two of the five problem statements, farm-level data is critical. As we will discuss in Section 3, primary data collection is the best, and often the only, way for IDH Farmfit to generate farm-level data of a sufficient quality. For instance, when assessing SDM effectiveness, farm-level data is collected to measure the improvement of farmer livelihoods in terms of increased incomes. Beyond income, we focus on the improved conditions of income creation, such as access to services, asset ownership, and gender dynamics (e.g., decision making at household and farm-level). To capture farmer resilience, we collect data related to the adaptive capacity of farmers, such as cash flows, savings, food security, and climate resilience (e.g., perception indicators on climate risks, climate-related crop loss, and related mitigation strategies).

In addition, farm-level data indirectly sheds light on how SDM operators can improve their efficiency and financial viability. Accurate farm-level data gives SDM operators a better understanding of types of services farmers need; how they can best be delivered; how access to better services can reduce risks and costs of service provision to SHFs and improve farm profitability; and how SHF engagement models can be adapted accordingly. Understanding the value that an SDM creates for farmers can also help SDM operators identify appropriate strategies for making their SDM more sustainable (e.g., by charging for specific services at appropriate price levels).

Across these problem statements, IDH Farmfit focuses on several impact areas, covering gender, climate resilience, and food security and nutrition (see Box 3 on how IDH Farmfit incorporates gender into its SDM analyses and primary farm-level data collection¹¹).

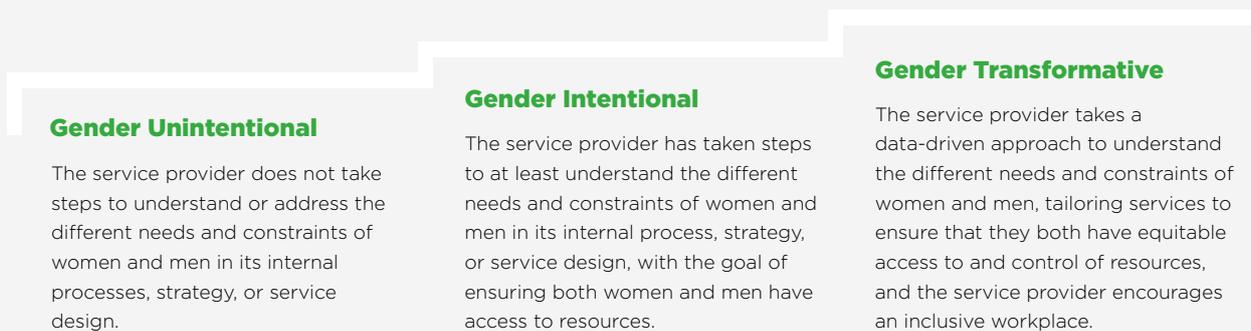
11. More information can be found in IDH Farmfit's recent publication, *Optimizing Farm Systems Through Gender Inclusion*, accessible via <https://www.idhsustainabletrade.com/gender-reports-optimizing-farm-systems-through-gender-inclusion/>

BOX 3 - THE IMPORTANCE OF PRIMARY DATA IN IDH FARMFIT'S GENDER RISK AND OPPORTUNITY ANALYSIS

IDH Farmfit's SDM methodology includes a gender risk and opportunity analysis to understand the degree to which an SDM operator incorporates gender into their SDM, and to identify opportunities for improving this for the benefit of both farmers and the SDM operator. The objective of the gender analysis is to help SDM operators become gender intentional or gender transformative, which are classifications referring to the extent that gender-specific strategies are integrated in SDMs (see Figure 4).

Applying a gender intentional or gender transformative approach may have additional cost implications for the SDM operator, both for initial design and for operationalization. These costs can be recouped as a result of increased levels of farm productivity, profitability, and customer loyalty from female farmers, all of which can improve the commercial viability and investability of the SDM. An overview of our gender risk and opportunity analysis—and insights from its application across SDM case studies—can be found in our recent publication, *Optimizing Farm Systems Through Gender Inclusion*.

FIGURE 4: STAGES OF GENDER INTENTIONALITY DEFINED BY IDH FARMFIT



Farm-level data—already lacking or of insufficient quality—is rarely disaggregated by gender. To apply a gender lens in our SDM analyses, primary data collection is crucial. The primary data we collect reveals the distinct needs and preferences of women and men and provides insights on how service provision can be made more valuable and effective for each gender-disaggregated segment. Primary data is thus critical for SDM operators to strengthen business and social outcomes for themselves, smallholder farmers, and broader value chains.

From the SDMs that we have analyzed using this gender lens, we have observed that female farmers still face gender-specific constraints that limit their productivity levels—for example, owning smaller plots and lower quality land; having less access to loans, as well as smaller loan sizes compared to male farmers; generating a more limited income; and having less decision-making power and autonomy on financial and farming matters.

The way forward

Despite pervasive gender gaps in agricultural value chains, IDH Farmfit believes that progress can be made if the following actionable steps are taken by SDM operators:

- To become “gender intentional,” SDM operators should at least understand the different needs and constraints of women and men. A first and critical step is commitment to the collection of gender-disaggregated data.
- To become “gender transformative,” the SDM operator should take a data-driven approach to tailoring services to ensure that women and men both have equitable access to, and control of, resources.
- In turn, IDH Farmfit will continue to support SDM operators with becoming more gender intentional and transformative, including in strengthening their data collection and analysis capabilities.

2.2

IDH FARMFIT’S DATA-DRIVEN METHODOLOGY TO IMPROVE SERVICE DELIVERY MODELS

Without solid data, IDH Farmfit would not be able to perform SDM analyses, design impactful technical assistance projects, or adequately apply its learning framework to achieve its mission. For each SDM analysis, three types of data are needed: 1) contextual data, 2) SDM-level data, and 3) farm-level data (see Table 1).

TABLE 1: THE TYPES OF DATA IDH FARMFIT COLLECTS AND USES IN A TYPICAL SDM ANALYSIS

Type of data	Examples (illustrative and non-exhaustive)
Contextual data	<ul style="list-style-type: none"> • Enabling environment (e.g., information on technology, environment, infrastructure, labor, institutional stability, land tenure, and social norms in the country/region) • Commodity value chain challenges in the country/region • Status of food security, gender, and climate resilience of farmers in the country/region • Status of living income of farmers in the country/region
SDM operator data	<p><i>SDM characteristics:</i></p> <ul style="list-style-type: none"> • Short- and long-term strategy and targets (operational, commercial, and financial) • Strategic objectives on farmer impact (such as farmer livelihoods, food security, gender, and climate resilience) • Business model (e.g., information on services, target farmer segments, and delivery channels) <p><i>SDM financial information:</i></p> <ul style="list-style-type: none"> • Overhead costs • Service-specific costs and revenues • Commercial sourcing data volume and margins (if SDM operator sources the commodity from farmers)
Farm-level data	<p><i>Farm characteristics:</i></p> <ul style="list-style-type: none"> • Geographical and crop-specific information, land size <p><i>Farm household characteristics:</i></p> <ul style="list-style-type: none"> • Gender ratio, food security and climate resilience status, educational level, land ownership, mobile phone use, loan accessibility, etc. <p><i>Agronomic information:</i></p> <ul style="list-style-type: none"> • Use and cost of labour, equipment, and inputs • Adoption of services • Farmer’s revenue data from main crop (e.g., yield, prices) • Farmer’s revenues from other crops, livestock, and/or off-farm income

Contextual data is used to provide an overview of external factors (such as the enabling environment) that the SDM operator may not influence but that affect the performance of the SDM. This data is important to understand how other actors can provide support—whether legislative, technical, or financial—to strengthen the SDM and maximize impact on farmer livelihoods. This data can typically be derived from secondary data, such as national statistics, academic literature, or expert input.

SDM operator data is used to provide an overview of the SDM (including the stakeholders and key objectives) and to quantify the costs and revenues associated with operating the SDM, both on an aggregate level and on the level of individual services. This data is used to evaluate historic and current SDM financials, as well as to perform a risk and opportunity analysis to create various projected scenarios for scaling the SDM. While operational and financial data is available in almost all cases from SDM operators, there may be a lack of clarity or reporting on the link between services provided and the impacts this can have on sourcing activities. This makes it difficult to identify—let alone quantify—the indirect benefits that an SDM can have: for instance, the effects on volume, security, quality, or efficiency of agricultural produce sourced from farmers reached through the SDM. Thus, in our SDM analyses, we also collect commercial sourcing data at the SDM level. Data at the SDM level is mostly obtained via SDM operators (i.e., through interviews, company monitoring and evaluation data, annual reports) and assumptions from local agronomists.

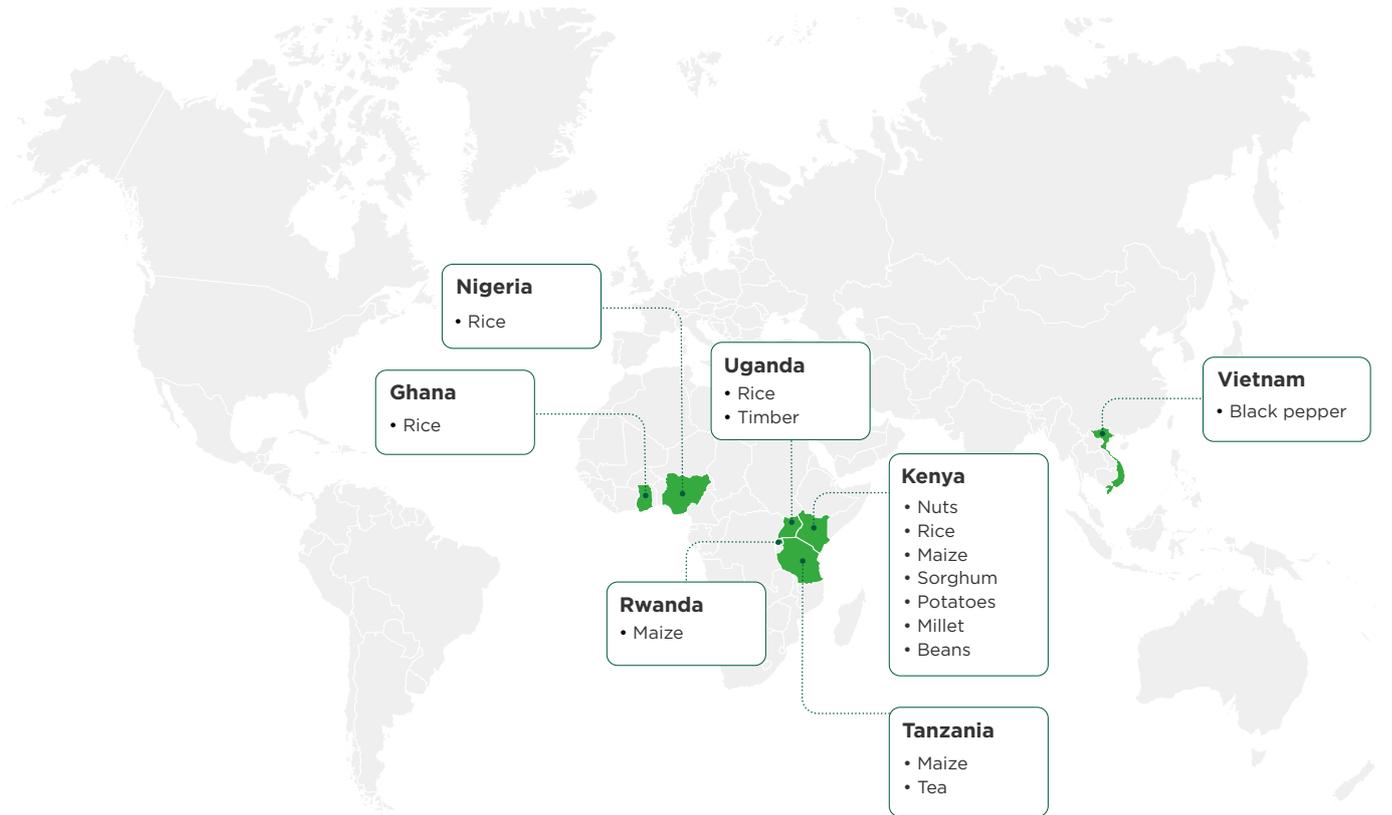
Farm-level data is used in our SDM analyses to assess performance risks and opportunities that farmers face, and to assess the value proposition of the SDM for farmers—specifically, to evaluate its impact on farm-level economics and social and environmental indicators. Given that SHFs are the primary customer of SDM operators accessing services—and, in many cases, may also supply agricultural products to SDM operators—a solid understanding of their socio-economic characteristics and agronomic behaviour is also crucial to designing and implementing sustainable, scalable, and investable SDMs. Without primary data, it is impossible



to get a comprehensive view of farmers to determine what needs, challenges, and opportunities exist at the farm level; the appropriate set of goods and services that should be provided to farmers; and how these impact farmer livelihoods and SDM performance.

Since 2019 IDH Farmfit has increasingly collected primary farm-level data and considers this to be core to our SDM analysis methodology and typically a pre-requisite for the provision of TA. To date, we have drawn insights from primary data collection as part of 18 SDM analyses across a variety of countries and value chains (see Figure 5). The number of farmers varies across each data collection, depending on the SDM and how this impacts the required sampling approach. On average, we collect data from 330 farmers for each SDM analysis.

FIGURE 5: PRIMARY DATA COLLECTED TO DATE BY COUNTRY/CROP



3.0

IDH FARMFIT'S KEY REASONS FOR PRIMARY FARM-LEVEL DATA COLLECTION

-
- 3.1 PRIMARY DATA STRENGTHENS SDM ANALYSES AND THE DESIGN OF TA PROJECTS**
 - 3.2 PRIMARY DATA GENERATES COMPARABLE AGGREGATE DATA FOR LEARNING ACROSS SDMS**
 - 3.3 PRIMARY DATA IS KEY TO EVALUATE THE IMPACT OF INTERVENTIONS ON FARMER LIVELIHOODS**

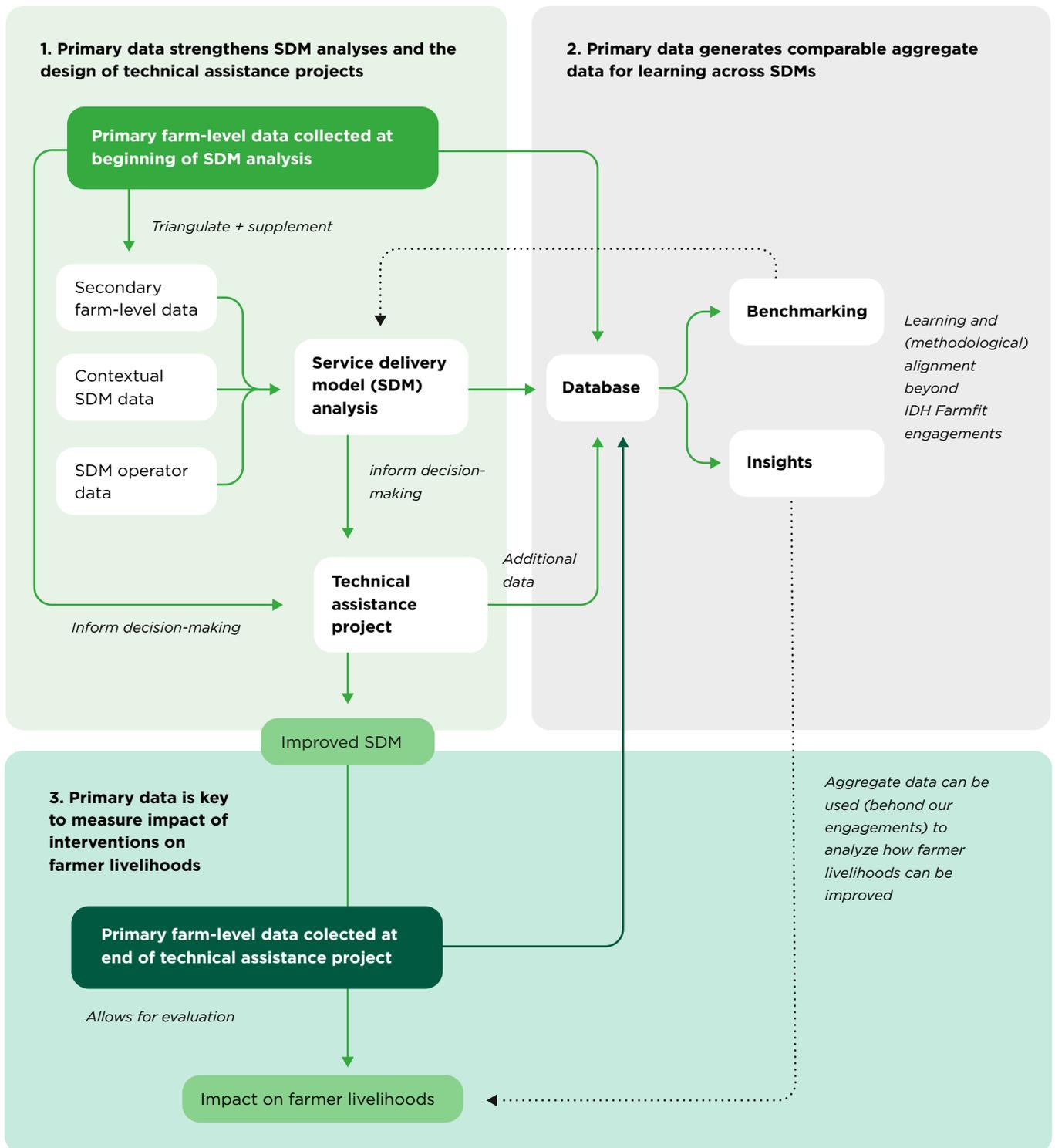


In this section we discuss how primary data collection is often the best way of collecting accurate data about SHFs. We also examine IDH Farmfit's key reasons for collecting this data, which are to:

1. Strengthen the analysis of SDMs and inform the design of TA projects with an aim of making SDMs efficient, effective, commercially viable, and investable at scale;
2. Generate comparable aggregate data that allows for comparison and learning across SDMs, enabling various stakeholders to make informed decisions and calculated risks on investments in smallholder farming; and
3. Measure and evaluate the impact of SDM operators' and our interventions on farmer livelihoods and adapt interventions accordingly.

The important role of primary farm-level data is demonstrated in Figure 6 and is further elaborated on the next page.

FIGURE 6: THE MULTIPLE OBJECTIVES OF PRIMARY FARM-LEVEL DATA FOR IDH FARMFIT



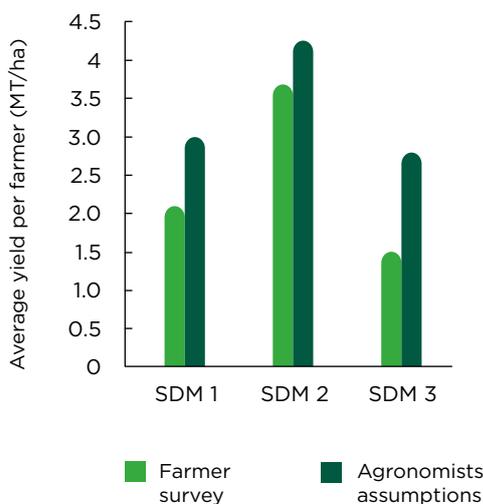
3.1

PRIMARY DATA STRENGTHENS SDM ANALYSES AND THE DESIGN OF TA PROJECTS

IDH Farmfit has observed and experienced numerous challenges related to the collection and usage of farm-level data. In some cases, the SDM operator has strong data available on the farmers they engage with, especially those companies where agronomists and field agents are in active contact with farmers and are supported by monitoring and evaluation systems. In some cases, SDM operators may operate detailed Farm Management and Information Systems (FMIS) or collect farmer field book data on a highly granular level. However, collecting such data for enough farmers over long enough periods of time and with adequate quality control is complex and costly. Many SDMs—particularly those run by small and medium-sized businesses—do not have the time, resources, or experience needed to invest in farm-level data collection, resulting in little or no data available or data that is not digitalized and hence difficult to analyse.

In our experience comparing data from agronomists and from primary data collection, it is common to find significant variation between these two sources. For instance, in three highly comparable SDM analyses the assumptions from SDM operators and local agronomists on annual production levels per farmer were higher than the annual production captured in the primary data for farmers (see Figure 7). Without solid data, it is often impossible to gain a credible understanding of the impact of SDMs on farmers. While primary data collection is often the best way of collecting accurate data, limitations may also exist, for instance when respondents’ answers are affected by recall bias or “social desirability” bias. Therefore, in each SDM analysis, we triangulate data from other sources—such as the SDM operators’ agronomists, existing research, and external experts—against our farm-level survey data. This added level of data improves the reliability of our dataset.

FIGURE 7: TRIANGULATION OF YIELD ASSUMPTIONS FROM AGRONOMIST AND FARMER SURVEY DATA



Primary data also enables us to collect relevant types of data needed to improve farmer livelihoods. For instance, it allows for farmer segmentation, by which farmers involved in the SDM are grouped into distinct segments, each of which represents a group of farmers with comparable characteristics, such as farm sizes, agronomic practices, and likelihood to adopt or use practices or inputs. This segmentation is important for understanding the diverse needs for and impacts of service provision on distinct groups of farmers. Segmentation allows IDH Farmfit and SDM operators to better understand how farmers each have distinct challenges and needs to which a well-designed SDM could be tailored (see Box 4).¹²

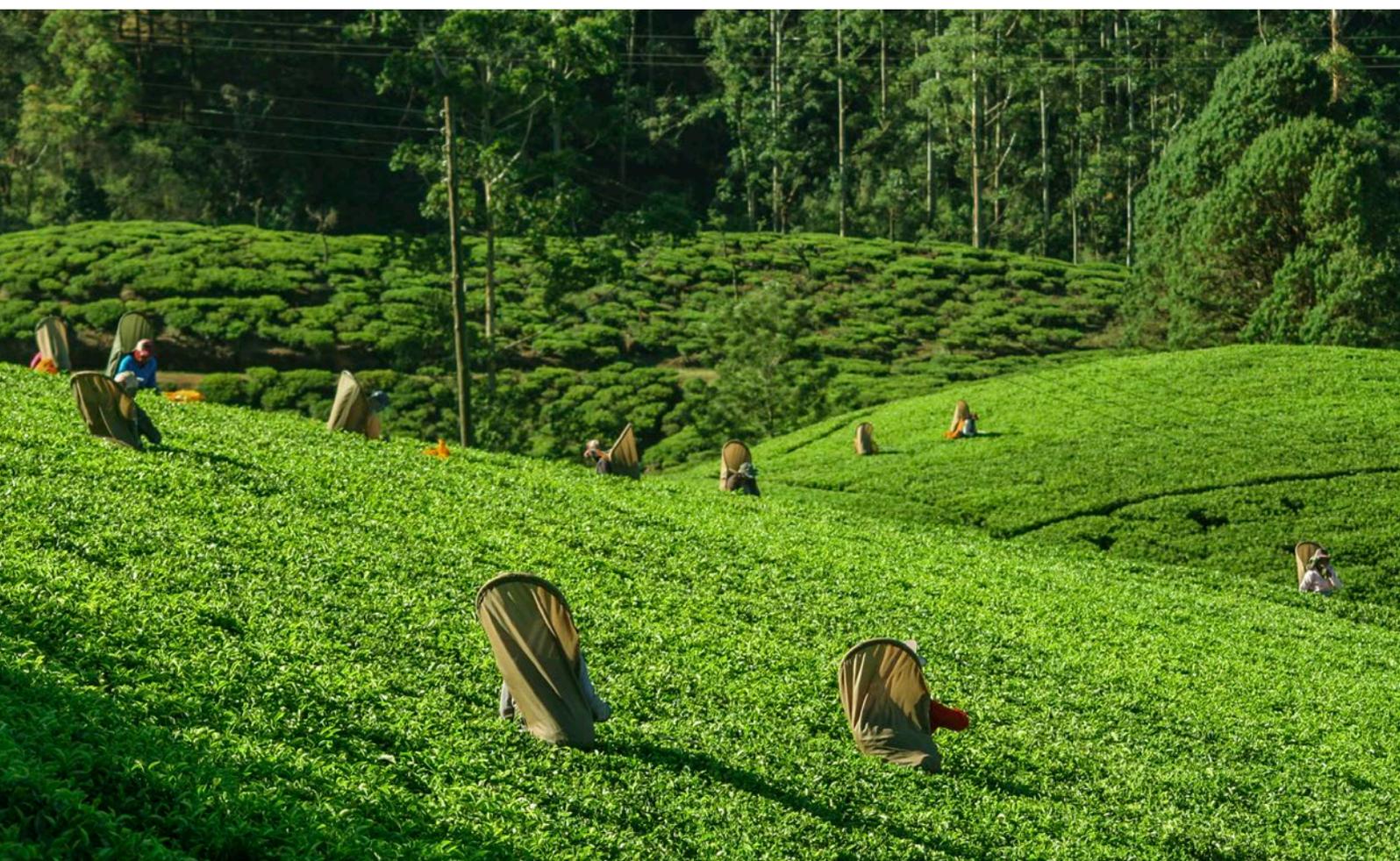
12. More information can be found in our case studies accessible via: www.farmfitintelligence.org/resources

BOX 4 - PRIMARY DATA USED TO GUIDE SEGMENTATION IN TEA SMALLHOLDER FARMING IN TANZANIA

Primary data was used as part of an engagement in which the SDM operator was a cooperative and minority shareholder in a tea processor representing approximately 15,000 smallholder tea farmers in Tanzania. The cooperative provided a wide range of services, such as training, tea bush seedlings and agro-inputs on credit to tea farmers to 1) secure and increase supply by raising farm productivity, and 2) improve farmer incomes from tea and other crops.

When IDH Farmfit began its engagement with this cooperative, limited data was available on the tea farmers. Primary data collected from the tea farmers as part of the SDM analysis was critical in informing farmer segmentation to better understand farmer characteristics and needs.

The SDM analysis showed that, for certain segments, more investments were needed in tea bush density (e.g., infilling with tea seedlings), whereas other segments needed to invest more in tea bush yield (e.g., good agriculture practices and access to fertilizers). The outcome of the analysis assisted the cooperative in designing the SDM to better achieve its objectives, specifically improving the productivity of tea farmers and their income. The farm-level primary data can also indirectly improve the performance of the SDM operator, as a more effective and impactful SDM for farmers is more likely to raise productivity and farmer loyalty, improving sourcing volumes and supply security for the SDM operator.



3.2

PRIMARY DATA GENERATES COMPARABLE AGGREGATE DATA FOR LEARNING ACROSS SDMS

The SDM methodology was explicitly designed in a way to produce transferable assets¹³ that can be shared and used externally. These transferable assets take the form of data and insights that we have collected and analysed in a standardized and comparable way. The volume and quality of primary data collected from SHFs for each SDM analysis and TA engagement contributes to IDH Farmfit’s large farm-level dataset, allowing for benchmarking.

Our benchmarks are used in different ways and can benefit various types of stakeholders. SDM operators that work with IDH Farmfit gain additional insights into their business model through comparison against, and inspiration from, benchmarks. A broader set of stakeholders—including our peers, as well as service providers, farmers, knowledge institutions, and governments—can benefit from access to aggregate data made available on our web portal¹⁴ and aggregate insights on best practices, lessons learned, and trends delivered through our knowledge products (i.e., publications, workshops, and conferences). Stakeholders can use these data and insights to make smarter decisions in smallholder agricultural markets. The insights that IDH Farmfit generates based on this ever-growing dataset pave the way for investors to make well-informed decisions and calculated risks on investments in smallholder farming.

“Comparable farm-level data creates an opportunity for benchmarking performance across programs. It facilitates a chance to learn best practices from other programs and avoid pitfalls that they have made along the way. With a large enough dataset, it is possible to develop performance benchmarks. What’s unique about these benchmarks is that they aren’t based on secondary data, which is often collected in a different context. They are based on the data collected from the actual recipients.”

– Reflection from 60 Decibels

However, to promote aggregate insights, we need data that is both accurate *and* comparable across different contexts, such as value chains and geographies. Directly collecting data allows us to use industry-aligned indicators and collate information in a standardized way to ensure comparability across SDMs (our approach is further elaborated in Section 4).¹⁵

13. Transferable Assets are packages of mature innovations, either public or private goods, which address system-level constraints and have a high potential for scaling across a wide range of environments. (definition adopted from the Bill and Melinda Gates Foundation)

14. Accessible via www.farmfitIntelligence.org/resources

15. We store all the primary farm-level data in a central database, which facilitates comparison across different surveys. This has enabled us, for example, to identify key cross-context trends in gender dynamics at farm-level in our recent publication, *Optimizing Farm Systems Through Gender Inclusion* (see Box 3). Accessible via <https://www.idhsustainabletrade.com/gender-reports-optimizing-farm-systems-through-gender-inclusion/>

We want to expand the impact beyond sharing data and insights. This report is part of that ambition; we hope that our methodology and lessons learned on farm-level primary data collection can benefit the wider community active in the smallholder agriculture space. By sharing our methodological approach to data collection, we seek to contribute to alignment in the industry—for example, operationalizing concepts into industry-relevant indicators and approaches to collecting and analysing data. In the longer run, our ambition is to convene other organizations in the sector to use increasingly harmonized language and methodologies, such that the data and insights generated in this manner will become increasingly interoperable. Through these efforts, the ease and value of benchmarking will be continuously strengthened.



“Mainstreaming data collection across projects, programmes, and investments allows for higher levels of reliability, but also for cross-case comparison and context analysis. The latter is important for purposes of scaling up/out and understanding within which context results are (potentially) applicable. The major challenge here is disagreement about metrics and methods of measurement: because there is too much inconsistency and incoherence these data cannot (easily) be merged to create larger data sets than can be analysed.”

- Reflection from Wageningen University & Research

“Having comparable data and data collection instruments would be a huge win. When data is comparable and shared in standard formats, it enables larger-scale data sharing and use that can support the development sector at large. It can support better policy making, targeting of resources, and programmatic design.

Having data and data collection instruments that are comparable would reduce the need for smaller actors to collect data. This would be a benefit for smaller actors, as data collection can be time consuming and relatively costly for them. If they could retrieve part of their data from credible databases they can spend their resources on only collecting the data that is crucial and still missing.”

- Reflection from Akvo Foundation

3.3

PRIMARY DATA IS KEY TO EVALUATE THE IMPACT OF INTERVENTIONS ON FARMER LIVELIHOODS

An important part of IDH Farmfit's mission—and that of our peers—is achieving positive impact on farmer livelihoods. In fact, we believe that SDMs' success at achieving positive impact at the farm level is the foundation for each model's sustainability, scalability, and investability. Good quality farm-level data, collected using primary data collection methods, is crucial to assessing whether an SDM has indeed realized positive impacts on farmer livelihoods. To this end, IDH Farmfit intends to carry out additional endline SDM analyses, including primary data collection after each TA project. The intention is not only to report what has changed at SDM and farm levels since our support, but also to evaluate and learn why certain outcomes have emerged.

In contrast to impact evaluations that rely on secondary data obtained via the SDM operator for farm-level impact, IDH Farmfit uses primary data for impact evaluation at farm level. We compare various indicators of farmer livelihoods before our engagement with private sector partners (i.e., the beginning of the SDM analysis and TA implementation) to the same indicators at the end of a TA project. To do this, we need primary data from farmers to have an accurate understanding of their livelihoods—especially since we include not only economic indicators, such as farmer income, but also social indicators, such as gender equality and resilience against climate shocks. We believe that such impact evaluation, based on credible farm-level data, is not only important for the understanding and decision making of our private sector partners, but also for our own decision making.





“In times that we become increasingly data driven, creating insights that drive our decisions, accessibility and reliability of data become more important than ever. Any project, programme, or larger-scale investment will require sound data, not only for purposes of rigorous impact assessment but also for guiding decision making and (re-) steering management.”

– Reflection from Wageningen University & Research

Although it is challenging to establish causal effects by comparing baseline and endline situations, we can nevertheless detect changes in farmer livelihoods that have occurred since our engagements, monitor our intended outcomes, and, if needed, adjust our interventions. Furthermore, during the course of a TA project we regularly collect qualitative data from SDM operators and farmers to further shed light on changes at the SDM and farmer levels in the context of the SDM and our interventions. Simultaneously, we are exploring the possibility of conducting randomized control trials (RCTs) in a TA project; RCTs are considered an effective tool to test causality (i.e., knowing that what is being achieved is a direct result of the intervention). This approach, which relies on primary data collection is a powerful way to identify cost-effective interventions for donors investing in evidence-based policy.

“High-quality data can help to make development programs more effective and efficient, it helps to target resources and programs. Having data directly from the farmer ensures you can target and monitor intended outcomes and contributions to impact.”

– Reflection from Akvo Foundation

16. ATAI (2019) *Evidence for transformation: framing a research agenda in agriculture for development*. Accessed via <https://www.atai-research.org/atai-program-renewal-framing-paper/>

4.0

IDH FARMFIT'S METHODOLOGY, CHALLENGES, AND BEST PRACTICES TO PRIMARY DATA COLLECTION

-
- 4.1 SCOPE OF THE DATA COLLECTION
 - 4.2 STANDARDIZATION AND MODULARITY
IN SURVEY
 - 4.3 ADAPTATION OF THE SURVEY TO THE
SDM AND LOCAL CONTEXT
 - 4.4 DATA COLLECTION PROCESS

To meet IDH Farmfit’s objectives and address the above-mentioned challenges to farm-level data collection, we have carefully developed a research design. This section provides an overview of our methodological approach to farm-level data collection, including best practices to overcoming challenges during data collection.

This methodology design is a result of continuous collaborative efforts by both IDH Farmfit and our data collection partner, Akvo Foundation (hereafter called Akvo), which executes and oversees the data collection process. Akvo has worked with governments and organisations in more than 70 countries in the field of agriculture to improve implementation and data-driven decision making of development projects.

4.1

SCOPE OF THE DATA COLLECTION

Farm-level data is collected through a structured farmer household survey.¹⁷ The design choices for setting the scope of data collection are reflected in Table 2. The unit of analysis within each SDM analysis is primarily the individual smallholder farmer¹⁸ in charge of the day-to-day management of the farm. Different members of the household respond to a number of survey questions. For instance, female decision making in the household and access to resources require the primary female decision maker in the household to respond, whether they are in charge of the farm or not. Additionally, questions related to food security are targeted at the household member who generally prepares the food for the household. Furthermore, various questions about the entire household are included—for instance, on non-farm activities and income sources—to get a better understanding of broader farmer livelihoods. These questions are addressed by the primary respondent, the person who is in charge of the farm.

The SDM analyses cut across different value chains and geographies, and the types of services examined may therefore differ depending on the context.

17. For the design of the survey, existing instruments and methods were considered, including: 1) Feed the Future Monitoring, Evaluation, and Learning Toolbox; 2) Living Standard Measurement Study - Integrated Surveys in Agriculture (World Bank); 3) Rural Household Multi-indicator Survey (RHOMIS); and 4) Living Income Community of Practice (LICO-P). The survey was further refined to address the data needs and research objectives of the Farmfit program.

18. Smallholder farmers are defined by the SDM operator, as the definition of what constitutes a smallholder farm varies per country and crop.



TABLE 2: SCOPE OF DATA COLLECTION PER SDM ANALYSIS

Boundaries of the survey	Description
Main unit of analysis	The individual farmer
Boundaries of the SDM	1-2 geographical areas (provinces, districts) and 1-2 focus crops
Sample size	On average, 330 farmers are surveyed per SDM analysis
Reference period	The data reflects a 12-month reference period
Duration survey	On average, 35 minutes is required to administer the survey per farmer

4.2

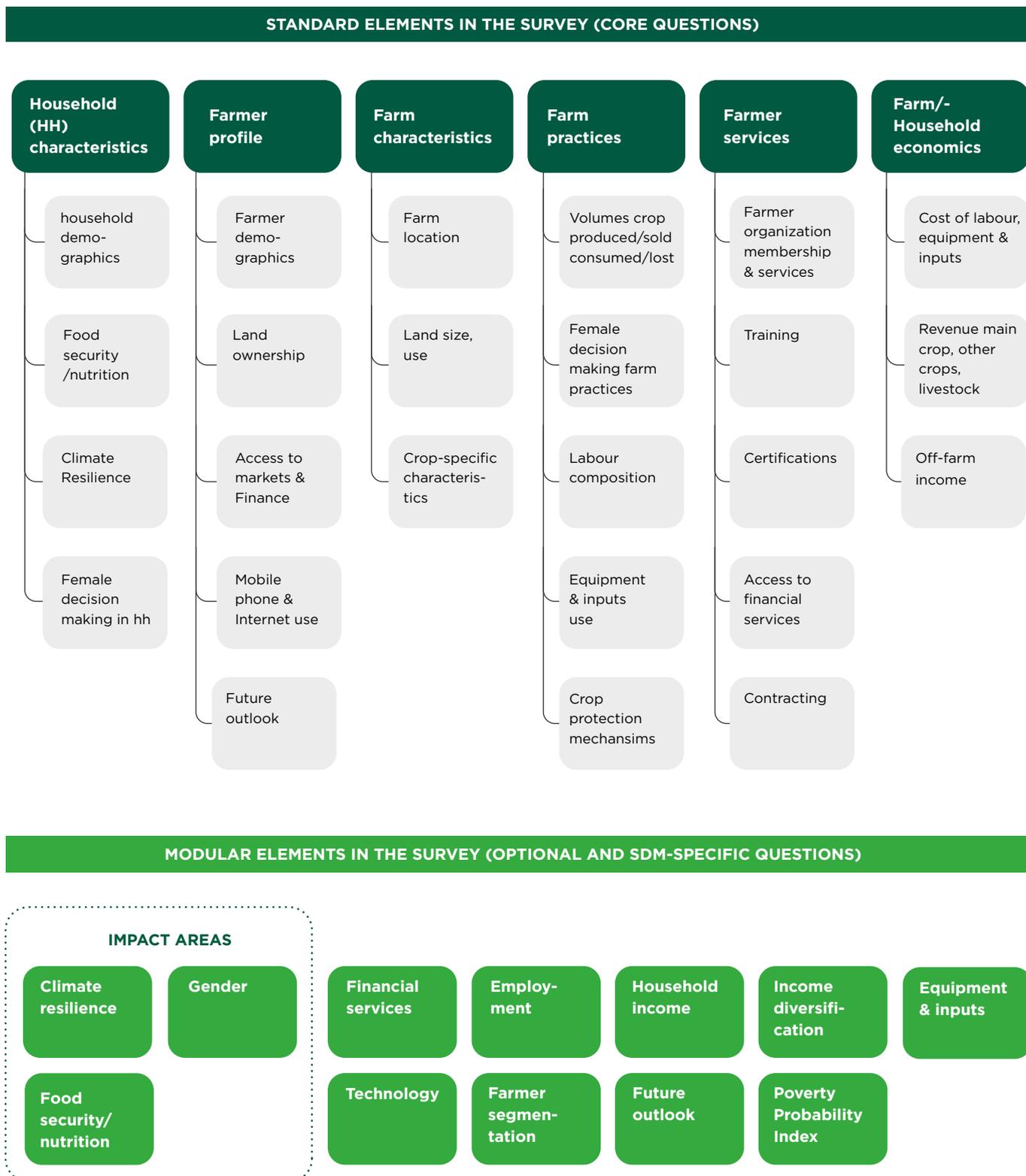
STANDARDIZATION AND MODULARITY IN SURVEY

As discussed in Section 2.2, our methodology for SDM analyses is meant to generate data from different SDMs that can be aggregated to derive comparable data. Therefore, it is critical to collect the data in a standardized way. Efforts are made to ensure a balance between standardization and modularity in the survey design. The survey is standardized in the sense that, for each data collection effort, the same questions are asked to ensure that answers can be compared across different SDMs. At the same time, the survey design also allows for modularity; it is possible to add questions specific to the SDM or to dive deeper into a certain problem statement or theme.

In collaboration with Akvo, IDH Farmfit has developed a Question Library to facilitate the survey design for each data collection in a rigorous, efficient, and consistent manner. The library consists of a comprehensive set of survey questions, including standard core questions, SDM-specific questions, and optional questions to dive deeper into certain themes. Figure 8 shows the different topics covered in the standard and modular elements of the survey.



FIGURE 8: QUESTION LIBRARY DEMONSTRATING TOPICS OF STANDARD AND MODULAR ELEMENTS IN THE SURVEY



4.3

ADAPTATION OF THE SURVEY TO THE SDM AND LOCAL CONTEXT

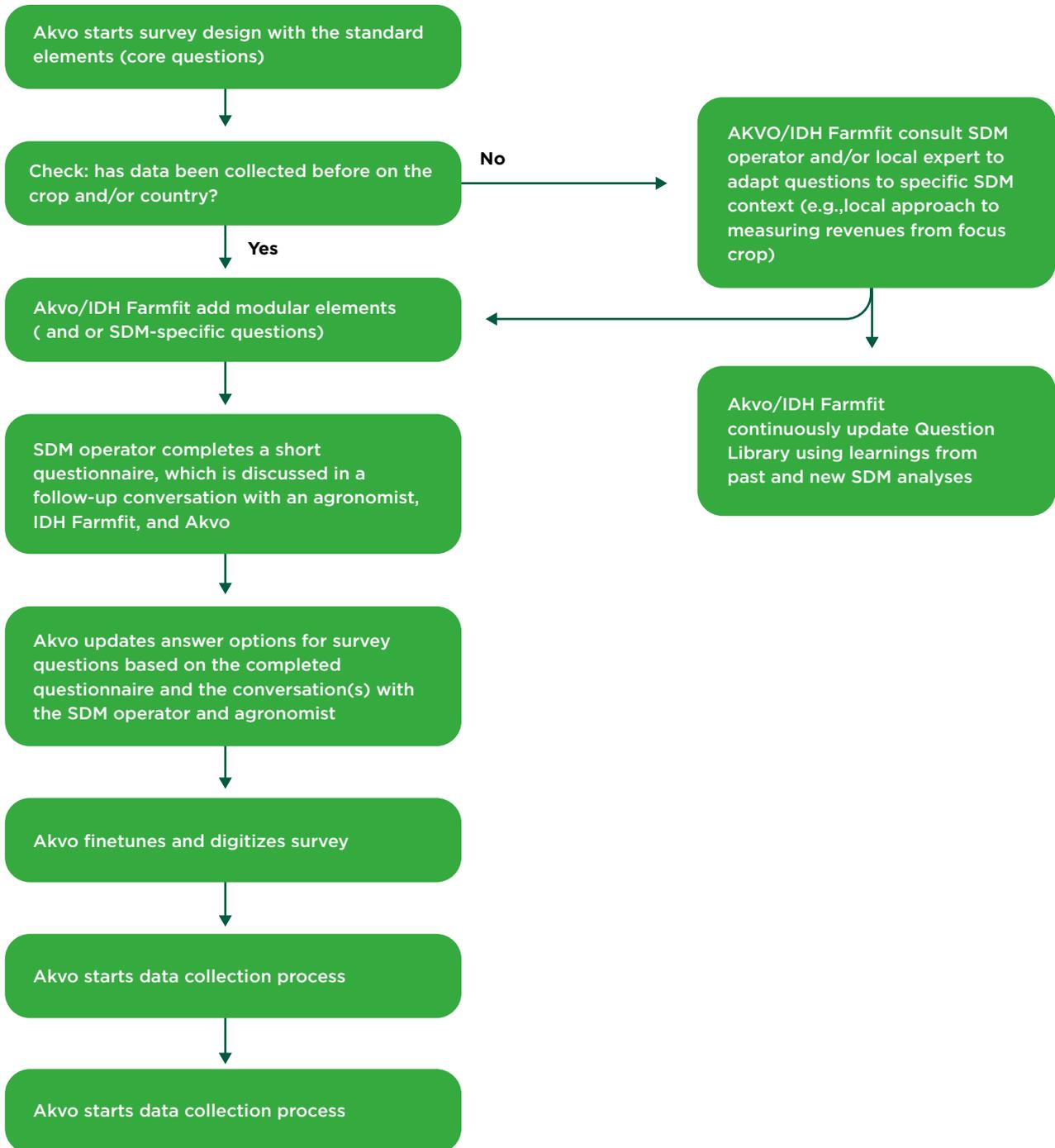
Figure 9 shows the various steps taken by the data collector (Akvo) and IDH Farmfit to continually assess and improve the data collection approach. Whenever there is a need for data collection for an SDM analysis where the country and/or focus crop is new and no primary data has been collected before by either IDH Farmfit or Akvo, both parties will familiarize themselves with the new context to tailor the survey questions accordingly. Although the survey is largely based on standardized questions, the SDM and its context may differ; hence some questions or answer options need to be adapted to capture such nuances. For instance, depending on the focus crop that is central to an SDM, measurement units for yield or other variables may differ (e.g., approaches to capturing revenue from sales in cocoa differ from those in coffee due to differences in selling practices). These contextual differences also need to be reflected in possible survey questions and answer options.

A critical step in this process is verifying information with the SDM operator and/or local agronomists to assess whether survey questions and answer options are properly defined and adapted to the given context. This is to improve data quality and minimize outliers, as well as to lessen the burden on farmers during data collection.¹⁹ The new, context-specific questions and answer options are then added to the Question Library, further enriching our data collection tool for future use.²⁰ This process, together with the vast experience of IDH Farmfit and Akvo in the field of smallholder agriculture, enhances accuracy of measurement and validity of our analyses. Through our localized and tailored data collection approach we address one of the key data challenges in agriculture, which is measurement complexity due to vast differences in crop, geography, and other local aspects.

19. Various other measures are taken in the survey design (e.g., skip logic, exhaustive answer options, double entering of numerical value, etc.) to maximize the quality of the data.

20. Our enriched database is also used as input by the data collector in optimizing the calculation of sampling strategies for future data collections.

FIGURE 9: CRITICAL STEPS IN THE SURVEY DESIGN



4.4

DATA COLLECTION PROCESS

RECRUITMENT AND TRAINING OF ENUMERATORS

IDH Farmfit has developed an extensive research and data collection plan together with Akvo. A period of eight weeks is calculated to ensure that Akvo can carefully execute this plan for a particular project. A key element is the selection and training of local enumerators. These are selected based on their experience of working in the agriculture sector; in the absence of such experience, enumerators are recruited from the nearest university that offers a specialization in agriculture. Additional selection criteria for enumerators include: prior experience with mobile-based data collection; background in agriculture; general understanding of financial concepts; fluency in local language/dialect; and familiarity with the geographical area of data collection. Recruiting local enumerators who can pick up local nuances and ask for clarification on the spot has proven to be key for getting accurate and informative responses from farmers.

During a training workshop the enumerators learn how to use the data collection application on a smartphone, familiarize themselves with the survey, practice interview techniques, and learn how to troubleshoot during field data collection.

DIGITAL DATA COLLECTION TO ENSURE DATA QUALITY

Throughout the data collection process, digital tools are used to ensure data quality. The enumerators visit farmers to administer the survey face-to-face while storing the data digitally (typically using smartphones). The survey process is adapted on an ongoing basis to improve data quality. For instance, to avoid numerical typos made by enumerators, questions were created that allow for double entry (i.e., the enumerator enters a response twice for numerical questions) and detect any wrong input, such as non-numerical input for numerical questions.

The incoming data is visualized in a data collection tracking dashboard that provides an overview of the exact location of data collection, the number of data points collected by each enumerator, the time spent surveying participants, whether the surveyed farmers were part of the actual sample, and how many farmers refused to participate. This is done to ensure that there is an audit trail of enumerators visiting the farmers. To mitigate data quality issues at an early stage, Akvo's data science team reviews the incoming data to detect any issues or patterns that may affect data quality. For example, the team looks for recurring mistakes by enumerators or missing values for certain questions. This dashboard also enables the identification of outliers. For instance, outliers on production and farm size are examined by monitoring data on the amount of focus crop produced compared to the size of land dedicated to the focus crop.



BOX 5 - DATA COLLECTION DURING COVID-19

At the time of writing, we are undertaking data collection during the global COVID-19 pandemic. Alternative measures have been taken to facilitate remote data collection when needed. Akvo has adapted its software to facilitate phone-based data collection and tested this approach during a pilot in the beginning of 2020. The approach proved to be effective in getting responses in contexts where in-person data collection is not advisable or possible. The following factors have

proven to aid phone-based data collection: 1) readily available phone numbers of respondents provided by the SDM operator; 2) restricting the length of the survey to avoid respondent fatigue; 3) scheduling an introductory call with respondents to decide on a date and time for the interview; 4) provision of monetary incentives, such as phone credits, to participate in the survey; and 5) oversampling to ensure the data is representative.

4.4

DATA COLLECTION PROCESS

RECRUITMENT AND TRAINING OF ENUMERATORS

IDH Farmfit has developed an extensive research and data collection plan together with Akvo. A period of eight weeks is calculated to ensure that Akvo can carefully execute this plan for a particular project. A key element is the selection and training of local enumerators. These are selected based on their experience of working in the agriculture sector; in the absence of such experience, enumerators are recruited from the nearest university that offers a specialization in agriculture. Additional selection criteria for enumerators include: prior experience with mobile-based data collection; background in agriculture; general understanding of financial concepts; fluency in local language/dialect; and familiarity with the geographical area of data collection. Recruiting local enumerators who can pick up local nuances and ask for clarification on the spot has proven to be key for getting accurate and informative responses from farmers.

During a training workshop the enumerators learn how to use the data collection application on a smartphone, familiarize themselves with the survey, practice interview techniques, and learn how to troubleshoot during field data collection.

DIGITAL DATA COLLECTION TO ENSURE DATA QUALITY

Throughout the data collection process, digital tools are used to ensure data quality. The enumerators visit farmers to administer the survey face-to-face while storing the data digitally (typically using smartphones). The survey process is adapted on an ongoing basis to improve data quality. For instance, to avoid numerical typos

21. Snowballing is a non-probability sampling technique used by Akvo and refers to the sampling of farmers based on referrals from the SDM operator and farmer group leads. Akvo selects farmers who receive services from, or sell their produce to, the SDM operator and ensures that the referrals are not based on neighbouring farms to increase the chance of collecting data from farmers with different characteristics and farming practices.

“Farmers who live in remote areas cannot always be easily reached or targeted. To ensure that samples will be representative, and give every farmer a chance to participate in the survey regardless of their location, there is need to put into consideration the farmers in the remote areas as early as possible in the design to ensure that there is proper time/budget allocated to visit those farmers.”

– Reflection from Akvo Foundation

made by enumerators, questions were created that allow for double entry (i.e., the enumerator enters a response twice for numerical questions) and detect any wrong input, such as non-numerical input for numerical questions.

The incoming data is visualized in a data collection tracking dashboard that provides an overview of the exact location of data collection, the number of data points collected by each enumerator, the time spent surveying participants, whether the surveyed farmers were part of the actual sample, and how many farmers refused to participate. This is done to ensure that there is an audit trail of enumerators visiting the farmers. To mitigate data quality issues at an early stage, Akvo’s data science team reviews the incoming data to detect any issues or patterns that may affect data quality. For example, the team looks for recurring mistakes by enumerators or missing values for certain questions. This dashboard also enables the identification of outliers. For instance, outliers on production and farm size are examined by monitoring data on the amount of focus crop produced compared to the size of land dedicated to the focus crop.

TIMELY EXECUTION OF THE DATA COLLECTION

As in many studies in the smallholder agriculture sector, IDH Farmfit has faced challenges related to the time and resources spent on data collection. Collecting the data at an early stage of the SDM analysis is critical; it allows the IDH team to guide the analysis and speed up the process of triangulating data sources. However, we have experienced

delays in the execution of primary data collection for several reasons; the main one is delays in contracting processes with the SDM company related to primary data collection. This underscores the importance of engaging the SDM client as early as possible in conversations on primary data collection, to demonstrate the value of primary data collection and portray it as a vital element of an SDM analysis. It’s also important to clarify the assistance required from the SDM operator in the collection process (e.g., assisting with permit



22. A set of protocols and procedures were developed, which are not limited to Farmfit but apply to the wider organization of IDH and need to be followed when IDH contracts an independent party. More information can be found via <https://www.idhsustainabletrade.com/privacy-statement/>

23. NDA's are also signed by the locally recruited enumerators who administer the surveys among farmers.

applications). Clear guiding principles for remote data collection have also been critical (see Box 5).

SAMPLING FRAME AND RESPONSE RATE

IDH Farmfit and Akvo depend on the SDM operator's assistance in providing a sampling frame (i.e., an overview of the farmers) from which a representative sample can be selected. In cases where SDM operators do not have good records of all the farmers they work with, there is a risk of a potential bias to the sampling strategy. SDM operators might

“Contrary to public health and related domains, there are few government regulations of economic, agricultural, and other related research. Where the former requires ethical codes and protocols to secure privacy of subjects, the latter are exempted from such practices. Therefore, respecting privacy seems to be a voluntary act and dependent on the ethical codes of the firm or organization.”

- Reflection from Wageningen University & Research

“It is important to also think about ways in which data can be used by farmers themselves. From a systems perspective, the observations from an individual farmer are often stored in a database that is not accessible to the farmers. The direct benefit that farmers can have from the learnings are therefore most likely not coming back. We do see solutions for this, considering increasing digitalization.

We think that data should always be accessible to local farmer groups. Farmers should ideally be able to learn from problems farmers in certain villages encounter, and possible solutions to address these. Putting the needs of farmers central in design is key, but while it sounds logical to do so, it is often not current practice.”

- Reflection from Akvo Foundation

5.0

CONCLUDING REMARKS

5.1 KEY REPORT TAKEAWAYS

5.2 NEXT STEPS IN THE ROAD TO
DATA-DRIVEN IMPACT AT FARM LEVEL

5.1

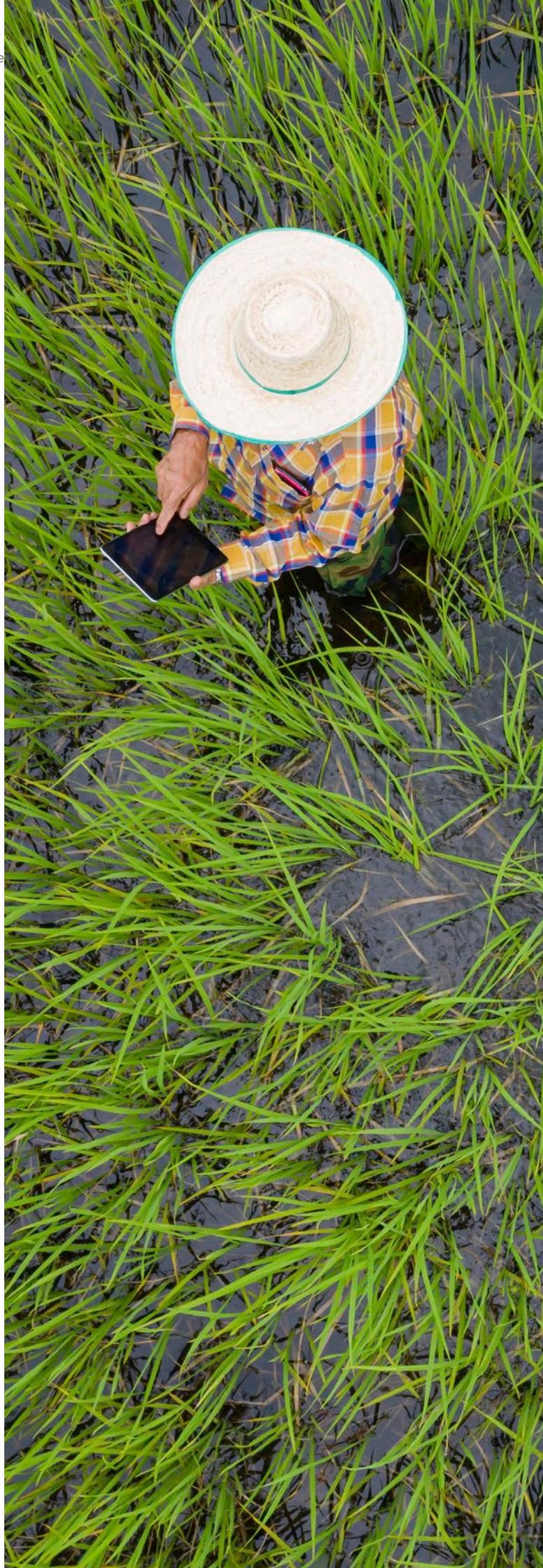
KEY REPORT TAKEAWAYS

In this report, we demonstrated that agricultural data—including farm-level data—is critical to create financially viable and investable service delivery models that improve farmer livelihoods. Therefore, a critical component of our private sector engagements is the collection of reliable, comparable, and interoperable farm-level data to achieve our mission of contributing to the transformation of smallholder agricultural services markets. More specifically, high quality farm-level data adds value to our work in the following ways:

1. Strengthen the analysis of SDMs and inform the design of TA projects with an aim of making SDMs efficient, effective, commercially viable, and investable at scale;
2. Generate comparable aggregate data for comparison and learning across SDMs, allowing various stakeholders to make informed decisions and calculated risks on investments in smallholder farming; and
3. Measure and evaluate the impact of interventions on farmer livelihoods critical to adapting interventions to maximize positive impact on farmer livelihoods.

Yet numerous challenges associated with farm-level data collection exist; many of which we have experienced within the IDH Farmfit program. We believe these can be best tackled through a well-developed research design and data collection process, as shared in this report. Furthermore, a challenge less easily overcome by a single actor or organization is the lack of shared learning and comparability of primary data across different organizations, which has also been emphasized by our knowledge partners (Akvo Foundation, 60Decibels, and Wageningen University & Research).

We believe that the road to data-driven impact for SHFs is best travelled together with our knowledge partners and industry peers. This involves engaging in conversations and collaborations related to data—such as the use of harmonized indicators, learning questions, collection methodologies, and data ownership and sharing mechanisms—to enhance shared learning and interoperability of data and insights.



5.2

NEXT STEPS IN THE ROAD TO DATA-DRIVEN IMPACT AT FARM LEVEL

By sharing our lessons learned in this report we hope to encourage the broader ecosystem of actors to adopt a data-informed approach to improving farmer livelihoods, and to provide practical guidance to those seeking to collect farm-level data. Simultaneously, we are keen on engaging with other actors (e.g., development organizations and knowledge partners) to collaborate on our data-driven journey, share insights, and create further alignment in the industry.

Specifically, we are interested to learn how others are dealing with: 1) farm-level data ownership and privacy; 2) how to effectively share back information to farmers; 3) ensuring representation of women in farm-level surveys; and 4) challenges and best practices regarding the use of technologies for data collection (e.g., Farmer Information Management Systems). In addition, we welcome engagement on methodological approaches to capture and measure concepts around farmer livelihoods, including living income,²⁴ gender inclusion, food security, and climate resilience.

Finally, the long-term sustainability of SDMs depends on attracting financing and investments, and this requires a solid understanding of the ability of SDMs to create sufficient and monetizable value for agribusinesses, SHFs, and other market stakeholders—for which farm-level data is crucial. Therefore, it is imperative for stakeholders to view primary data collection not as an add-on, but as critical to the longevity of inclusive agricultural business models. We recognize that such data collection requires resources and sparks various questions, such as: What is the future role of primary data in smallholder agricultural value chains? How does this fit within the increasing adoption of Farmer Management Information Systems? And who is to take responsibility for funding data collection? We aim to prove the value of farm-level primary data and hope that, in the future, data collection (and costs associated with it) will be internalized in the business models of SDMs.

“60 Decibels works with investors, corporates, foundations, NGOs, and social enterprises. A reasonable proportion of our work is funded by private sector actors. What they find most compelling is that they receive actionable insights within a matter of weeks, which allows them to improve their operations. For too long impact measurement/primary data collection has been pitched as a compliance requirement. But, at its essence, it is a chance to learn and improve. As long as the tools being used make that possible, private sector players are willing to invest in them.”

- Reflection from 60 Decibels

“Data is the new gold. While the development sector still treats it largely as a ‘cost,’ the private sector, and particularly the IT sector, sees it as a valuable good. Investments in data can help to better develop products and services, and we see the large corporates all moving into this space with massive investments. So while the current users are mainly public and development sector actors, the playing field is rapidly changing—and we would argue that the development sector at large is often behind in adopting proper data management procedures and policies, particularly those that can help to ensure the rights of end users.”

- Reflection from Akvo Foundation

24. IDH has recently published the “Income Measurement Survey,” which captures most elements of farmers’ actual income, and has been built upon the Question Library discussed in Section 4. More information can be accessed via <https://www.idhsustainabletrade.com/news/idh-supports-companies-in-taking-action-towards-closing-the-living-income-gap>

IDH Farmfit continues to take next steps on the road to impactful service delivery to smallholder farmers, including:

- Refining primary data collection within our own operations to generate better data for decision making and improve the impact IDH Farmfit has through SDM analyses, technical assistance, funding, and dissemination of insights to the broader sector;
- Supporting private sector partners with building and investing in better data collection and entrenching this in their SDMs and exploring ways to feed data and insights back to SHFs; and
- Collaborating with peers in conversations related to data—such as data ownership, the use of harmonized indicators, learning questions, and collection methodologies—to enhance shared learning and interoperability of data and insights.



**Farmfit
Intelligence**

**Get in
Touch!**

Charlotte Keijser
IDH Farmfit Intelligence
keijser@idhtrade.org