



Platform for Big Data in Agriculture

Towards a core approach for cross-sectional farm household survey data collection: a tiered setup for quantifying key farm and livelihood indicators

Mark van Wijk, Cristina Alvarez, Guvvala Anupama, Elizabeth Arnaud, Carlo Azzarri, Dharani Burra, Francesco Caracciolo, David Coomes, Alessandra Garbero, Elisabetta Gotor, Jessica Heckert, Nancy Johnson, Soonho Kim, Berta Miro, Jacqueline Muliro, Kelvin Shikuku, Marcelo Tyszler, Roberto Valdivia, Sara Viviani, Hans Vrolijk, Gideon Kruseman



001
100
111
001

101 100
000 100
110 100
000 111
111 1011

000 110 01
010 01 11
100 010 100
100 10 10
010 10 10

10100 10010 10
001110 10 1110
111 1000 0010
011 110 000

001 010 00 110
00001 1010 11 111
1001100110100 00001 100 101 100
0110 11101010110011000 00000 100100

00111 0010110101110000 11 00 100
0 11011 101110001 11110
110 101001000 101 1110
01110 10110111101100 101 110
11100100 101101110011 10 110000
001 100100110 10001100 1010001

0100 1101000110100 110 100
101 10011101000 1110
110100 11100 00110
100011000 000 100
11011100 110 000

1010 100 10
010 11 11 0
000 100 01 0
100 100 00 100
0001101 011100



Towards a core approach for cross-sectional farm household survey data collection: a tiered setup for quantifying key farm and livelihood indicators

Mark van Wijk, Cristina Alvarez, Guvvala Anupama, Elizabeth Arnaud, Carlo Azzarri, Dharani Burra, Francesco Caracciolo, David Coomes, Alessandra Garbero, Elisabetta Gotor, Jessica Heckert, Nancy Johnson, Soonho Kim, Berta Miro, Jacqueline Muliro, Kelvin Shikuku, Marcelo Tyszler, Roberto Valdivia, Sara Viviani, Hans Vrolijk, Gideon Kruseman



About the authors

Mark van Wijk¹, Cristina Alvarez¹⁶, Guvvala Anupama², Elizabeth Arnaud³, Carlo Azzarri⁴, Dharani Burra⁵, Francesco Caracciolo⁶, David Coomes⁷, Alessandra Garbero¹⁶, Elisabetta Gotor³, Jessica Heckert⁴, Nancy Johnson⁸, Soonho Kim⁴, Berta Miro⁹, Jacqueline Muliro¹⁰, Kelvin Shikuku¹⁰, Marcelo Tyszler¹¹, Roberto Valdivia¹², Sara Viviani¹⁵, Hans Vrolijk¹³, Gideon Kruseman¹⁴

¹ ILRI (International Livestock Research Institute, Nairobi), Kenya

² ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), Patancheru, Hyderabad, India

³ Bioersity International, Montpellier / Rome, France / Italy

⁴ IFPRI (International Food Policy Research Institute), Washington DC, USA

⁵ CIAT (International Center of Tropical Agriculture), Hanoi, Vietnam

⁶ University of Naples, Naples, Italy

⁷ Evans School of Public Policy and Governance, University of Washington, USA

⁸ CGIAR Independent Science and Partnership Council, Rome, Italy

⁹ IRRI (International Rice Research Institute), Los Baños, Philippines

¹⁰ Worldfish, Penang, Malaysia

¹¹ KIT (Royal Institute of the Tropics), Amsterdam, the Netherlands

¹² Oregon State University, Corvallis, USA

¹³ Wageningen University, Social Sciences Group, the Hague, the Netherlands

¹⁴ CIMMYT (International Maize and Wheat Improvement Center), Texcoco, Mexico

¹⁵ FAO, Rome, Italy

¹⁶ IFAD, Rome, Italy



About the CGIAR Platform for Big Data in Agriculture

The CGIAR Platform for Big data in Agriculture aims at using big data to solve agricultural development problems faster, better and at greater scale. Data has become a valuable global commodity, but it is much more than simply information: in expert hands, it is intelligence.

Already, analysts are finding ways to turn big data – the immense stocks of information collected in computers worldwide – into an invaluable resource for planning and decision-making. It is helping accelerate the development of robust responses to some of the most pressing challenges of our time: climate change/variability, food insecurity and malnutrition, and environmental degradation. The smart and effective use of data will be one of the most important tools for achieving the United Nations' Sustainable Development Goals. Big data represents an unprecedented opportunity to find new ways of reducing hunger and poverty, by applying data-driven solutions to ongoing research for development impact.

About CoP_SED

The Community of Practice on socio-economic data (SED-CoP), led by CIMMYT, aims at bringing together CGIAR centers, academia, not-for-profit research and development organizations and private sector partners willing to tackle major issues related to socio-economic data.

The community works together on strategies to make the data interoperable, in order to enhance the impact and the use of CGIAR-related socio-economic data for partners in development.

This space can be used as a discussion area, share and request relevant information and contribute towards building the community as a whole.

Citation

Mark van Wijk, Cristina Alvarez, Guvvala Anupama, Elizabeth Arnaud, Carlo Azzarri, Dharani Burra, Francesco Caracciolo, David Coomes, Alessandra Garbero, Elisabetta Gotor, Jessica Heckert, Nancy Johnson, Talip Kilic, Soonho Kim, Berta Miro, Harriet Mugera, Jacqueline Muliro, Kelvin Shikuku, Marcelo Tyszler, Roberto Valdivia, Sara Viviani, Hans Vrolijk, Gideon Kruseman. 2019. *Towards a core approach for cross-sectional farm household survey data collection: a tiered setup for quantifying key farm and livelihood indicators*. Community of Practice on Socio-economic Data report COPSED-2019-001. CGIAR Platform for Big Data in Agriculture.

License

Creative Commons Attribution License 4.0 International
<https://creativecommons.org/licenses/by/4.0/legalcode>



Acknowledgements

We gratefully acknowledge the contributions of Talip Kilic, Harriet Mugeru of the World Bank group and Flavio Bolliger of FAO at and around the meeting on “data standardization and harmonization” held at IFAD, Rome December 5-7, 2018. We are very grateful to IFAD, Rome for hosting the workshop.

This work was funded through mini-grants of the Community of Practice on socio-economic data of the CGIAR Platform for Big Data in Agriculture. There were in kind contributions from: ILRI (International Livestock Research Institute, Nairobi), Kenya; ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), Patancheru, Hyderabad, India; Bioversity International, Montpellier / Rome, France / Italy; IFPRI (International Food Policy Research Institute), Washington DC, USA; CIAT (International Center of Tropical Agriculture), Hanoi, Vietnam; University of Naples, Naples, Italy; Evans School of Public Policy and Governance, University of Washington, USA; CGIAR Independent Science and Partnership Council, Rome, Italy; IRRI (International Rice Research Institute), Los Baños, Philippines; Worldfish, Penang, Malaysia; KIT (Royal Institute of the Tropics), Amsterdam, the Netherlands; Oregon State University, Corvallis, USA; Wageningen University, Social Sciences Group, the Hague, the Netherlands; CIMMYT (International Maize and Wheat Improvement Center), Texcoco, Mexico; FAO, Rome, Italy; IFAD, Rome, Italy.

Foreword

In May 2017, the CGIAR Platform for Big data in Agriculture was officially launched. Around that time, the nascent community of practice on socio-economic data held its first virtual meet-ups. One of the first things that emerged from these discussions was the recognition that any form of standardization in the socio-economic domain was conspicuously absent. In an effort to redress the challenges facing socio-economic data reuse and data interoperability, three working groups emerged. The working group 100Q focused on identifying key indicators and related questions that are commonly used and could be used as a standard approach to ensure data sets are comparable over time and space. The working group SociO! focused on the development of a socio-economic ontology with accepted standardized terms to be used in controlled vocabularies linked to socio-economic data sets. The working group OIMS focused on the development of a flexible and extensible, ontology-agnostic, human-intelligible and machine-readable metadata schema to accompany socio-economic data sets.

This report is the first from the community of practice on socio-economic data and presents the results of the work conducted in the working group 100Q.

September, 2019

Gideon Kruseman, community of practice on socio-economic data coordinator



Table of Contents

About the CGIAR Platform for Big Data in Agriculture	iii
About CoP_SED	iii
Citation	iii
License	iii
Acknowledgements	iv
Foreword	iv
1. Introduction	1
1.1. Key aspects covered	2
2. Household composition and characteristics	3
3. Farm characteristics	4
3.1. Land available and use	4
3.2. Livestock use and production	8
4. Income and assets	12
5. Gender	16
6. Food security and dietary diversity	19
7. Other aspects	22
8. Overall Conclusions and Recommendations	23
References	24
Annex 1 List of attendants of the workshop Rome, Italy, in December 2018	25



1. Introduction

There is an urgent need to improve the characterisation of agricultural systems at household level to enable a more efficient assessment of the capacity households to adopt a range of agricultural intervention options. Local drivers and factors need to be identified that might constrain or provide opportunities within a specified agricultural system (Carletto et al., 2015), while on the other hand generalisable standardized characteristics need to be identified that would allow robust comparisons between different systems (Frelat et al., 2016; van Wijk et al., 2014).

The assessment of opportunities at smallholder farm household level to improve their livelihoods needs integration of validated standardised agricultural, poverty, nutrition and gender indicators in the quantitative characterisation of these households. This will allow us to assess how these welfare indicators vary across a farm household population and across different agro-ecological and socioeconomic conditions. Such data would also allow us to better assess how they may change over time.

Furthering such a standardization across all institutes within the CGIAR (who have been estimated to conduct baseline interviews with around 180,000 farmers per year) would allow for much easier application of big data method applications for analyzing the household level data themselves, as well as for linking these data to other larger scale information sources like spatial crop yield data, climate data, market access data, roadmap data, etc. The Big Data platform of the CGIAR has therefore stimulated an effort to define how a common core of a cross-sectional household survey focusing on rural households could look like, the so-called 100Q exercise (with 100Q standing for 100 Questions that that core should contain). The core survey should deliver key information around the agricultural activities and off farm income of the household, as well as key welfare indicators focusing on poverty, food security, dietary diversity and gender equity.

Within this effort a workshop was held in Rome, Italy, in December 2018, where a group of scientists from different centers of the CGIAR and partner institutions discussed how such a core approach for cross-sectional surveys could look, and what type of information should be captured. This report is a short reflection of what was discussed during this workshop, and tries to summarize the overall conclusions of this workshop into core modules of key aspects and indicators of rural farm livelihoods. This information can be used as building blocks for survey development, thereby resulting in more harmonized household survey data collection across CGIAR centers.



1.1. Key aspects covered

Based on evaluation of a wide range of survey instruments (e.g. Frelat et al. 2016; RHoMIS, Hammond et al., 2017 and WorldBank's LSMS-ISA survey tool (Living Standards Measurement Study – Integrated Survey on Agriculture) we decided that the key aspects that should be present in the core version of a household survey are the following:

Household composition and characteristics (Chapter 2)

Farm characteristics (Chapter 3)

- *Land available and use*
- *Livestock available and use*

Income and assets (Chapter 4)

Gender (Chapter 5)

Food security and Dietary Diversity (Chapter 6)

Other aspects (Chapter 7)

The information underpinning these will cover key information underpinning several of the Sustainable Development Goals (SDGs), and then in particular SDG 1 (no poverty), 2 (zero hunger), 5 (gender equity) and 12 (consumption and production).

In the following chapters we will dive into how each module for these aspects could look like, and what simple and more elaborate approaches can be within these modules (i.e. different 'Tiers'). This report deals with questionnaire design, and therefore does not deal with questions related to the survey application (e.g. to whom do you ask the questions, what are appropriate sampling strategies, etcetera).



2. Household composition and characteristics

This is a basic module of all farm household survey questionnaires, and it is not really possible to distinguish different Tiers here. Key variables to be collected are the age, sex and education level of all household members. The main difficulty of this module actually lies in the definition of a 'household' and which members belong to a household and which members do not. For example, members of the rural household can live in the city and send money home. So they are part of the livelihood but not necessarily of the household as they do not help with managing the farm nor do they eat from 'the same pot'. Other complicating factors for the definition of a household are the well-known family clusters that occur in West Africa where extended families live together and for some aspects (for example food or livestock herd management) function as a single unit whereas for money and crop management often smaller family units make decisions, or the occurrence of polygamy where a single husband supports more than one family.

So often two aspects are combined to define the household and its members in households: 'who is eating from the same pot at least three months a year' and 'who are making decisions about and managing their agricultural resources'.

The household roster can then be asked for like in Table 1, following the setup used in the LSMS-ISA survey.

Table 1: Example of a household roster table with the key variables of interest

Household member nr.	Age [years]	Sex	Relation to household head ¹	Main occupation ²	Number of months resident in house during past 12 months	Education level ³
1						
2						
3						
4						
5						
Etc.						

¹ Head; Spouse; Son/daughter; Son/daughter in law; Grandchild; Parent or parent in law; Other relative (male or female); Other (not related)

² Farming; Salaried worker, Self-employed; Student, Retired/not able to work

³ With education levels as: illiterate; primary; secondary; higher education



3. Farm characteristics

For farm characteristics we want quantitative information about the agricultural production resources (land & livestock), their key crops and species, and the main use of the produce.

Logically this information can be divided into two subsections, 1 – Land available and use, and 2 – Livestock available and use

3.1. Land available and use

Basic key information to be collected here are how much land is under cropping ('land cultivated'), how much land is owned, and whether common land resources are used as well. Different variations are possible, and within-year variations between cropping seasons are likely to occur. An overview of what such a module could look like is given in Table 2. Note that this information is based on farmer recall, and that this might lead to biases in the land area estimations. The 'gold standard' approach would be to GPS the fields of the farms surveyed to assess the reliability of the recall information. We realize, though, that that for many projects might be a bridge too far.

Table 2: A basic set of questions regarding land availability

Starting question (M = Multiple options possible)		
Does your h/h own land, rent land, use common land (for growing crops, fish culture or grow-out, or grazing animals)? (M)	<ol style="list-style-type: none"> 1. Own land 2. Rent in land for own use 3. Rent out land to others 4. Use common land 5. No, don't use any land 	
Question	Value	Unit
What was the total amount of land used by your household for growing crops last year?		
What is the total amount of land used by your household for fish culture or grow-out?		
How much land does your h/h own personally?		
Does your h/h have a kitchen garden or other place where you grow vegetables and fruits for home consumption?		

A widely used more detailed approach ('Tier 2') is to go beyond this basic information and ask for individual fields (or parcels), their size and their main land use. An example of such an approach is



given in Table 3, based on the LSMS-ISA way of asking this information (where LSMS-ISA takes a seasonal approach as the survey is conducted for each cropping season).

Table 3: A more detailed approach at field/parcel level to get insight into land allocation

Parcel Nr.	How large is the land area of [Parcel Nr.] that your household use?	Unit	Does [PARCEL] belong to your household? ¹	How did your household normally use the land area in [Parcel Nr.] during last few seasons? ²
1				
2				
3				
Etc				

¹ 1 Yes, entirely; 2 Yes, communal; 3 No, we rent it from others; 4 No, we sharecrop 5 No, we borrow at no cost

² 1 Annual crops; 2 Seasonal crops; 3 Tree crops; 4 Livestock; 5 Wood lots; 6 Other;

This field level approach can then also be the basis to ask for field differentiated management information, like mineral fertilizer application, manure application and irrigation. This would then also allow for more differentiated information on input costs, normally a weak point in many household and farm surveys. In the Tier 1 approach this can only be asked at overall farm level and later asked for the different crops (see below), and not at individual field level which limits the analyses focusing on land management that are possible. Although the field level approach might seem the logical choice for land use, it is important to note that it involves a simplification of a farm into consistent units of land management that is not applicable in all farming systems. Fields can change from season to season and from year to year. Parts of fields can be managed in certain ways, other parts in other ways. In complex, multi-cropping based systems with sometimes 8-9 crops in varying densities across a single piece of land, the field approach might not shed much light on land use. These limitations need to be taken into consideration when deciding on a certain approach.

Crops and their production

A basic approach here is to ask for the crops, fruits and vegetables cultivated on the farm and to ask for their main usage: for consumption, sales or for both. Table 4 below is an example of how such a set of questions can be laid out. This information will give a quick overview of the plant production diversity on farm and the farm orientation.



Table 4: Basic information regarding crop, vegetable and fruit production

Crops	Main use (consumption, sales, both)
Vegetables	Main use (consumption, sales, both)
Fruits	Main use (consumption, sales, both)

A more detailed approach ('Tier 2') is to zoom in further and also ask for the production of the most important crops, fruits and vegetables (to ask for all crops, fruits and vegetables is in most cases too time consuming for such a generic survey as we are aiming for). The simplest version of this approach can be to only ask for production details of the most important crop, or let the respondent to decide what the 'most important crops' are and ask for the production details of these crops.

Crop production can be asked for at different levels of detail in terms of aggregation by field and in time (i.e. across multiple seasons). Each of these approaches has pros and cons.

The main advantage of the 'by crop' approach is speed and in some cases accuracy. These two advantages happen when there are many individual plots, with the same crop (say maize) in multiple plots. Asking questions about each of these individual plots can be time consuming, and in quite a number of cases the farmer does not know what the crop production of each individual plot was, but only knows the total amount of production. The time gained by applying the 'by crop' approach can run up to 30-40m, estimated by trial runs with the RHoMIS tool, which is a substantial amount of time. The main disadvantage of the 'by crop' approach is that 'crop by plot' management information cannot be asked for and therefore is lost.



Table 5. Crop production and use information asked for 'by crop'

'Out of all of the crops grown by your household in the last 12 months, please select the most important (quantity/bringing in food or money) and complete the following questions'

Crop name	Total Crop Production	Unit	During the past 12 months, was the crop harvest good, normal, or bad?	Did you grow this crop intercropped with other plants?	About how much of your land did you use for growing this crop during the last 12 months?	Main use (consumption, sales, both)	What proportion of the crop yield was used for consumption?	If you sold crop production, at which unit price was it sold?	Unit of price information

Please use key below to answer the next questions:

1 = All or nearly all (90-100%) 3 = About half of it (40-60%) 5 = A small amount (1-10%)

2 = More than half of it (60-90%) 4 = Less than half of it (10-40%) 6 = None (0%)

This same information can be asked at parcel level, at annual or at, more logical if you want to do that level of detail, seasonal level. An example of this approach (which goes a bit beyond we would consider to be the 'core' approach for farm household surveys) is the CCAFS ImpactLITE survey (Rufino et al. 2013).

Land Management and Agricultural Inputs

Please answer the following questions with a list of crops, or if applicable, write 'None'



Table 6. Basic land management questions

On which crops did you use fertilisers during the last 12 months?	
How much fertiliser in total was used during the last 12 months? (specify units)	
What types of fertiliser does your h/h normally use?	
On which crops did you use manures or compost during the last 12 months?	
On which crops did you use pesticides during the last 12 months? Include herbicides, fungicides and similar chemicals.	
Which crops did you irrigate during the last 12 months?	
For which crops did you use improved seed varieties during the last 12 months?	

3.2. Livestock use and production

The most basic approach for livestock is to ask for the ownership of different livestock species, and how many of each of these the household owns. Although this seems a simple question, getting reliable information is not always easy. Problems can arise especially in pastoral communities where farmers are often reluctant to give accurate absolute numbers of their cattle herd size (it is often seen as sensitive information in those communities). Another problem can arise in communities where cattle are shared across families, and cattle can be herded by one family (who can keep the production) while the livestock is owned by another family.

Table 7. Basic questions regarding livestock owned

Livestock species owned	Quantity

For more detailed information on the breeds of the different livestock species a simple follow up question can be formulated as in Table XXX.



Table 8. Simple follow question on the livestock owned to get insight into the breeds of the different species.

Please select up to five MOST IMPORTANT livestock that you own and answer the following questions.

Insert livestock name	1	2	3	4	5
What breed are they?					
Local					
Cross-bred or exotic					
Mix (some local, some cross or exotic)					

Livestock production and use

The most basic livestock production and income generation information can be separated in two aspects: sales of live animals and consumption of livestock products and their sales. In this setup we ignore the slaughtering of animals and the consumption and/or sales of the meat produced. In some agricultural systems this can be an important use of animals, and in that case a similar setup as for livestock products can be used to gain quantitative information about its importance for the farm livelihood.

Table 9. Livestock production and use

Livestock product name	Total Production over the last year	Unit	Main use (consumption, sales, both)	What proportion of the production was used for home consumption?	If you sold livestock production, at which unit price was it sold?	Unit of price information
Cattle milk						
Goat milk						
Eggs						



Table 10. Basic information on live sales of the different livestock species present on a farm

Livestock name	Number sold in last year	Unit of price information
Cattle		
Goats		
Sheep		
Chickens		

Fish production and use

Another key activity of smallholder livelihoods is fish production or catching fish from open water bodies. To get insight into the importance of these activities the following series of questions can be used. The initial question

1. Did your household collect any fish or other aquatic animals in the last 12 months?

1-yes

0-no

Can, if answered with 'yes', be followed by more detailed questions regarding the quantity of fish caught, produced, consumed and sold, see Table 11.



Table 11. Key information regarding fish production and use. 'Indicate the amount collected from each source in its peak and lowest months:'

	Amount collected/produced during peak months	Unit	Amount collected/produced during low months	Unit	What proportion of the production was used for consumption?	What proportion of the production was sold?	Total cash value from sale of fish
<i>From own culture/grow-out stock</i>							
<i>From wild stock on household owned/rented property</i>							
<i>From wild stock off household property</i>							

To get insight into the variation of fish use over time, and possible catastrophic events in recent years, this section can be closed off with the question:

'During the last 3 years, did you experience any major (abnormal) mortality of fish that affected your production, incomes and livelihood?'



4. Income and assets

Agricultural income will be quantified using the sales information of agricultural produce collected above. This is gross income information (or rather value of production) as no cost information is included in its estimates. As noted before in the crops section, cost estimates are often a weak element in this type of farm and household surveys. We did not include a detailed set of questions regarding costs of input use in crop and livestock production (e.g. mineral fertilizer, irrigation costs, medicine use and purchase of fodder for livestock), because such a series of questions would take use beyond the ~100 questions envisaged in this exercise. It is however important to note that this is a limitation of the current set of questions identified in this report.

Key item to cover in this section is off-farm income. In many surveys off-farm income is asked for in absolute terms ('How much off farm income did you earn last year'), but there are clear indications that this leads to under-reporting by interviewees (e.g. Fraval et al., 2019). Another approach, used in for example the RHoMIS approach, is to ask for the relative importance of off farm income in the overall livelihood. By combining this information with the farm-based income one can quantify off farm income as well. Results indicate that this leads to lower under-reporting, with increased overall reliability of the survey results (Fraval et al., 2019).

Table 12. An example of how basic information on off-farm income can be collected

Does your household have any sources of income apart from selling what you produce on the farm?	Yes/No
Which types of off-farm income do you have? (Multiple option)	<ul style="list-style-type: none"> - Labour on other farms - Labour, not on a farm - Work in local business - Have own business - Remittances (send money) - Work for government or public institution - Rent out land to others
Consider all the money earned in the last 12 months from selling farm produce, and from the cash activities we just discussed. Did more money come from sales of farm produce, or from the off farm cash activities? ¹	

¹ 1 Almost All from farm; 2 Most from farm; 3 Half from off-farm; 4 Most from off-farm; 5 All or almost all from off-farm



Collecting information on spending can be approached in a similar way through the relative assessment way of asking as in Table 12. Costs are difficult to capture in single survey application, without diving deeply into individual activities and their associated costs. This is beyond the scope of this report. The most basic information that can be collected on spending is given in Table 13. This basic set of questions, which gives insight into whether earnings are being re-invested into the farm or mostly spent on the livelihood, can be expanded upon to get either a relative importance of each of the spending categories, or even can be asked for in absolute terms (but the latter with all the problems associated to asking questions about money in absolute terms in a one visit survey). The more detail that can be collected in terms of production costs (e.g. seed costs, labor and other inputs (fertilizer, pesticides, livestock vaccines and medicines) the better, as this information is crucial for many economic based impact assessments assessing the impacts of new/alternative technologies.

Table 13. Basic qualitative information on the spending allocation of farm households

		Income source:	Off-farm	On-farm
What does your household spend your earnings on? Answer for both off-farm and on-farm incomes. (M)	Buying food			
	Buying possessions (clothes, household items, vehicles)			
	Improving the farm (livestock, fertilisers, crops, machines)			
	Spend on people (education, health care, travel)			

Assets

Many different asset ownership approaches exist, which normally require a thorough effort to make them locally specific and relevant. One generic approach that can be used is the Probability of Poverty Index (PPI), which uses a set of 10 questions together with a scoring system that is calibrated on the poverty estimates derived from country level representative Living Standard Measurement Studies executed by the World Bank. Applying the scoring system to the answers given in a survey application results in an overall score that can be translated into a probability of that family being below or above the poverty threshold (now the 1.9 US\$ per person per day threshold). The PPI approach has attracted a substantial amount of criticism as a simplified and slow indicator of poverty, while also the PPI scoring system for some countries is quite old, and out-dated as it is calibrated on the old 1.25 US\$ per person per day threshold and not one the newer 1.9 US\$ threshold.



Table 14. The Probability of Poverty Questions for Tanzania

Indicators	Responses	Score
1. How many household members are 18-years old or younger?	A. Six or more	0
	B. Five	2
	C. Four	5
	D. Three	11
	E. Two	14
	F. One	17
	G. None	28
2. Are all household members ages 6 to 18 currently in school?	A. No	0
	B. Yes	3
	C. No members ages 6 to 18	5
3. What is the main building material used for the walls of the main building?	A. Baked bricks	0
	B. Poles and mud, grass, sun-dried bricks, or other	6
	C. Stones, cement bricks, or timber	13
4. What is the main building material used for the roof of the main building?	A. Grass/leaves, mud and leaves, or other	0
	B. Iron sheets, tiles, concrete, or asbestos	6
5. What is the main fuel used for cooking?	A. Firewood, coal, solar, gas (biogas), wood/farm residuals, or animal residuals	0
	B. Charcoal, paraffin, gas (industrial), electricity, generator/private source, or other	9
6. Does your household have any televisions?	A. No	0
	B. Yes	15
7. Does your household have any radios, cassette/tape recorders, or hi-fi systems?	A. No	0
	B. Yes	4
8. Does your household have any lanterns?	A. No	0
	B. Yes	4
9. Does your household have any tables?	A. No	0
	B. Yes	4
10. If the household cultivated any crops in the last 12 months, does it currently own any bulls, cows, steers, heifers, male calves, female calves, or oxen?	A. No crops, and no cattle	0
	B. No crops, and cattle	0
	C. Crops, but no cattle	5
	D. Crops, and cattle	12



However, there is no other generically applicable asset system available at the moment, which is why we chose to incorporate it in this report. We hereby give the example PPI for Tanzania in Table 14.



5. Gender

Insight into gendered information on asset ownership, decision power, voice and empowerment is typical information that cannot be gathered in one single module, but has to be gathered throughout a survey. The most basic information to perform sex-disaggregated analysis is already available through the household composition information (age, sex and education level of each individual household member), but key information can also easily be collected at several other moments in the survey

Asset ownership

Sex-differentiated ownership information of the productive resources available in the farm household is important information to get insight into asset ownership. When discussing how much land and livestock is owned, the question can be asked:

Who owns the land? With options (from which more than one can be chosen) available according to:

- Senior man of household
- Senior woman of household
- Male Child or Youth
- Female Child or Youth
- Other family member (male)
- Other family member (female)

Who owns the (cattle/goats/sheep/chickens/etc)? With the same options as above.

Table 15 and 16 give an overview how such expanded land and livestock tables look like



Table 15. Basic land use questions, including sex-disaggregated information (based on Table 2)

Starting question (M = Multiple options possible)		
Does your h/h own land, rent land, use common land (for growing crops or grazing animals)? (M)	<ol style="list-style-type: none"> 1. Own land 2. Rent in land for own use 3. Rent out land to others 4. Use common land 5. No, don't use any land 	
Question	Value	Unit
What was the total amount of land used by your household for growing crops last year?		
How much land does your h/h own personally?		
Does your h/h have a kitchen garden or other place where you grow vegetables and fruits for home consumption?		
Who owns the land? ¹		

¹ 1-Senior man of household; 2-Senior woman of household; 3-Male Child or Youth; 4-Female Child or Youth; 5-Other family member (male); 6-Other family member (female). Multiple options possible.

Table 16. Basic questions regarding livestock owned, expanded with sex disaggregated ownership information

Livestock species owned	Quantity	Who owns the livestock of this species? ¹

¹ 1-Senior man of household; 2-Senior woman of household; 3-Male Child or Youth; 4-Female Child or Youth; 5-Other family member (male); 6-Other family member (female). Multiple options possible.

Decision control

More detailed knowledge about the gendered differences in decision power over the benefits over on and off farm activities can be asked for in the crop, livestock and off farm income modules.

For each crop and livestock products one can ask the following two questions:



1. Who decides about how the farm produce is used? With options as above
2. Who decides about how the income generated by selling this product will be used? With options as above.

For off farm income one can also ask question number two, on who decides on how to use the money generated by off farm income. These questions can be added as extra columns to Tables 5, 9 and 10, and as extra row to Table 12.

This information allows for detailed sex disaggregated analyses on who can decide on how agricultural produce and resulting money is used, per individual crop, livestock and off farm income activity and overall at farm household level (see Tavenner et al., 2019 for such a detailed analysis).

Empowerment

To measure levels of empowerment in agriculture, IFPRI and partners have developed the Women's Empowerment in Agriculture Index (<http://www.ifpri.org/publication/womens-empowerment-agriculture-index>). This tool asks questions separately to men and women of the household (each interview taking about 45m) and generates empowerment scores for each. The scope of WEAI is beyond the simple core survey setup we aim for in this report, but the structured setup around 5 components of gendered information, i.e.

1. decisions about agricultural production,
2. access to and decisionmaking power over productive resources,
3. control over use of income,
4. leadership in the community, and
5. time use

allows for powerful in-depth analyses and across-site analyses.



6. Food security and dietary diversity

In recent years much work has been devoted to the harmonization of food security and dietary diversity indicators. Recently the FIES (Food Insecurity Experience Scale; <http://www.fao.org/in-action/voices-of-the-hungry/fies/en/>) emerged as a new standard for rapid food security assessments, and was also included as a core indicator for SDG2, Hunger. It consists of eight questions which can be combined into one score which can be calibrated on other food security status information, or used as a continuous scale by itself, or answers to individual questions can be used to classify households into moderately or severely food insecure. The FIES is seen as a further developed and targeted version of the HFIAS (Hunger and Food Insecurity Access Scale) developed within the USAID funded FANTA project.

The FIES questions are:

Think back over the last MONTH. Was there a time when, because of lack of money or other resources you personally...

1. Were worried you would not have enough to eat?
2. Were unable to eat healthy and nutritious food?
3. You ate only a few kinds of foods?
4. You had to skip a meal?
5. You ate less than you thought you should?
6. Your household ran out of food?
7. You were hungry but did not eat?
8. You went without eating for a whole day?

A second standard that has emerged in recent years to assess nutritional status is the MDD-W indicator (Minimum Dietary Diversity for Women). This is a simplified version of IDDS (Individual Dietary Diversity Score) in which just the women in the reproductive age category (normally 15 – 49 years of age) are targeted. For the official guidelines see <http://www.fao.org/3/a-i5486e.pdf>. Where indicators like Household Dietary Diversity Score (HDDS) assess the potential availability of food to the household, the MDD-W tries to go one step further to assess the real access of vulnerable groups within the household to food. The indicator also has a clearly defined threshold (5 out of 10) to determine whether a diet is adequate or not. MDD-W makes use of ten functional food groups (which are aggregated from the information on 16 food groups):

1. Grains, white roots and tubers, and plantains
2. Pulses (beans, peas and lentils)
3. Nuts and seeds
4. Dairy



5. Meat, poultry and fish
6. Eggs
7. Dark green leafy vegetables
8. Other vitamin A-rich fruits and vegetables
9. Other vegetables
10. Other fruits

These food groups in MDD-W are based on the functional use of food groups, and can therefore be used to approximate nutritional status. MDD-W can be asked for in open form, where in the interview the interviewee lists all items eaten during the last 24h, and the enumerator has to classify that information according to the food groups of MDD-W. In the closed form the enumerator asks directly whether a certain food group was consumed by listing example foods that are relevant for the region where the survey is executed and asking whether one or more of these food items were consumed over the last 24h (see Table 17).

The food groups that make up the MDD-W are mutually exclusive – that is, no food or ingredient is placed in more than one food group. Note that on the model questionnaire, three of the ten groups are further subdivided. This is for ease of recording and to make the questionnaire more intuitive for enumerators. For example, the food group “Meat, poultry and fish” is recorded on three rows (subgroups) on the questionnaire. Note that for more detailed micro-nutrient work especially the question on fish and seafood could be expanded with a question on the species of fish/seafood consumed. The micro-nutrient content of fish/seafood species is known to be very different.



Table 17. Did you consume these items over the last 24 hours?

A <i>Foods made from grains</i> Porridge, bread, rice, pasta/noodles or other foods made from grains	Yes	No
B <i>White roots and tubers and plantains</i> White potatoes, white yams, manioc/cassava/yucca, cocoyam, taro or any other foods made from white-fleshed roots or tubers, or plantains	Yes	No
C <i>Pulses (beans, peas and lentils)</i> Mature beans or peas (fresh or dried seed), lentils or bean/pea products, including hummus, tofu and tempeh	Yes	No
D <i>Nuts and seeds</i> Any tree nut, groundnut/peanut or certain seeds, or nut/seed “butters” or pastes	Yes	No
E <i>Milk and milk products</i> Milk, cheese, yoghurt or other milk products but NOT including butter, ice cream, cream or sour cream	Yes	No
F <i>Organ meat</i> Liver, kidney, heart or other organ meats or blood-based foods, including from wild game	Yes	No
G <i>Meat and poultry</i> Beef, pork, lamb, goat, rabbit, wild game meat, chicken, duck or other bird	Yes	No
H <i>Fish and seafood</i> Fresh or dried fish, shellfish or seafood	Yes	No
I <i>Eggs</i> Eggs from poultry or any other bird	Yes	No
J <i>Dark green leafy vegetables</i> List examples of any medium-to-dark green leafy vegetables, including wild/foraged leaves	Yes	No
K <i>Vitamin A-rich vegetables, roots and tubers</i> Pumpkin, carrots, squash or sweet potatoes that are yellow or orange inside	Yes	No
L <i>Vitamin A-rich fruits</i> Ripe mango, ripe papaya	Yes	No
M <i>Other vegetables</i> List examples of any other vegetables	Yes	No
N <i>Other fruits</i> List examples of any other fruits	Yes	No

Drawbacks to FIES and MDD-W

Both FIES and MDD-W (as well as many other implementations of dietary diversity, for example the Household Dietary Diversity Score) are typically based on recent recall information. The implementation period of the survey (before crop harvest, after crop harvest, at the end of the dry season) can therefore strongly affect the results that are obtained for these indicators. Another approach is therefore to ask for dietary or food security information for a specific period of the year (this is for example done in RHoMIS), thereby making the results independent of survey timing. However, recall length might make such an approach less reliable. The jury is still out on which of the two problems is most important (i.e. the within year variation of MDD-W or HDDS indicators or the recall error introduced by asking about a specific month earlier in the year), and new research is underway to assess both error terms.



7. Other aspects

This is an extra section where we want to capture key information related to the agricultural system and common entry points for agricultural interventions. We therefore ask for extra information regarding extension services, whether the farmer has tried new technologies recently and whether the farmer is part of a social network. All are seen as important drivers of whether farmers want to trial / adopt new technologies or other interventions. The questions are based on WorldBank's LSMS ISA surveys.

Extension services

Have you (= the household) received advice/information on vegetable gardens, crops, livestock, or soil and natural resource management in the last 12 months? 1 Yes 2 No

If 'Yes', from whom did receive this advice/information (Multiple answers possible):

1. Friend/neighbor
2. Model farmer
3. Other farmer
4. Farmer's group
5. Agricultural development/ extension agent

Innovation

Have you tried any new agricultural technologies/management practices during ...

last farming season 1 Yes 2 No

last two years 1 Yes 2 No

last three years 1 Yes 2 No

This question can be followed (if answered 'yes') by a question regarding the type of technology/management that was tried. For example was the technology crop/soil/livestock related? This can be expanded at will.

Network

Are you (= the household) a member of your community farmer or any other social organization or group? 1 Yes 2 No

And again this question can be expanded on by asking further (if answered 'yes') on which type of group they are member of. This, more detailed, question is likely to be locally specific.



8. Overall Conclusions and Recommendations

This report presented a simple, core approach that can easily be incorporated in household survey questionnaires. The questions presented here also form the basis of the mini-RHoMIS that has been developed recently and already applied by IITA in Rwanda and Burundi. In this way a powerful consistent and harmonized across site and across center database of household survey data can be constructed. A similar harmonized approach has been followed in the work of Frelat et al. (2016), Wichern et al. (2018) and Hammond et al. (2017), in which a common core of crop and livestock production information together with off farm income and household composition has led to a 40,000+ database of farm household data and associated indicators in 20+ countries in sub Saharan Africa (Waha et al. 2018; Van Wijk et al. 2018). By making such a data core publicly available (e.g. Van Wijk et al. forthcoming) new, insightful analyses in the pathways towards poverty reduction and increased food security in smallholder livelihoods can be performed at scale, beyond the single in-depth studies that are often performed. Furthermore, linking this information with other sources of (geo-spatial) data using a standardized library of concepts and variables (the current focus of the Big Data Socio-Economics Ontology workgroup) unlocks a new powerful world of analyses and continuous re-use of data for now and for in the future.



References

Carletto C, Jolliffe D, Banerjee R. 2015. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *The Journal of Development Studies* 51, 133-148.

Fraval S, Hammond J, Wichern J, Oosting SJ, de Boer IJM, Teufel N, Lannerstad M, Waha K, Pagella T, Rosenstock TS, Giller KE, Herrero M, Harris D, van Wijk MT. 2019. Making the most of imperfect data: a critical evaluation of standard information collected in cross-sectional farm household surveys. *Experimental Agriculture* 55 (2), 230-250

Frelat R, Lopez-Ridaura S, Giller KE, Herrero M, Douchamps S, Djurfeldt A, Erenstein O, Henderson B, Kassie M, Paul B, Rigolot C, Ritzema RS, Rodriguez D, van Asten P, van Wijk MT. 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proceedings of the National Academy of Sciences of the United States of America*, 113(2), 458-463.

Hammond J, Fraval S, van Etten J, Suchini JG, Mercado L, Pagella T, Frelat R, Lannerstad M, Douchamps S, Teufel N, Valbuena D, van Wijk MT. 2017. The Rural Household Multi-Indicator Survey (RHoMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agricultural Systems* 151, 225-233.

Rufino MC, Quiros C, Boureima M, Desta S, Douchamps S, Herrero M, Kiplimo J, Lamissa D, Mango J, Moussa AS, Naab J, Ndour Y, Sayula G, Silvestri S, Singh D, Teufel N, Wanyama I. 2013. Developing Generic Tools for Characterizing Agricultural Systems for Climate and Global Change Studies (IMPACTlite – Phase 2). Nairobi: ILRI.

WorldBank. (n.d) Living Standards Measurement Study – Integrated Surveys on Agriculture. Retrieved September 15, 2018 from <http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA>.



Annex 1 List of attendants of the workshop Rome, Italy, in December 2018

Carlo Azzarri (IFPRI)
Berta Miro (IRRI)
Hans Vrolijk (WUR)
Marcelo Tyszler (KIT)
Jessica Heckert (IFPRI)
Nancy Johnson (CGIAR)
Roberto Valdivia (Oregon State)
Francesco Caracciolo (University of Naples)
Elisabetta Gotor (Bioversity International)
Dharani Burra (CIAT)
David Coomes (University of Washington)
Guvvala Anupama (ICRISAT)
Elizabeth Arnaud (Bioversity International)
Soonho Kim (IFPRI)
Gideon Kruseman (CIMMYT)
Mark van Wijk (ILRI)
Talip Kilic (Worldbank)
Harriet Kasidi Mugeru (Worldbank)
Flavio Bolliger (FAO)
Cristina Alvarez (FAO)
Sara Viviani (FAO)
Alessandra Garbero (IFAD)

