INVENTORY OF DONOR MAPPING ON SDG2 RESULTS REPORTING

Final report

Ben Haagsma

15 November, 2018

Contents

1. Introduction

1.1 Background

The progress of SDG2 results has been the topic of a number of informal gatherings of donor agencies in 2017. Because the results of SDG2 were running behind on schedule, it was decided in January 2018 to improve coordination and generate more attention for SDG2. To that end a roadmap 2030 will be elaborated and presented on a SDG2 event the end of 2018. One element of this roadmap 2020 is a donor mapping that will present an overview on how bigger donor agencies report, in an aggregated way, on the results achieved by their funding. The indicators they use, should as much as possible connect or overlap with the agreed upon SDG2 indicators.

The client of this donor mapping was DGIS, Paul van de Logt, and it was assigned to Fair & Sustainable Consulting, the Netherlands, and done by Ben Haagsma. The mapping was implemented in the period July – October, 2018.

1.2 Purpose of the inventory

The purpose of this specific inventory or research is to contribute to better understanding and insights in global progress with regard to SDG2 results. The research questions were:

1) Which result areas and indicators (at different result levels) do the funding agencies report on in relation to SDG2 targets and indicators?
2) To what extent are these different result areas and indicators comparable and aggregable? Looking at the aggregated picture at different result levels.
3) How could donor agencies improve their planning and reporting on results and contributions to SDG2 targets in a more meaningful way? These contributions must be plausibly attributable with implemented projects and/or programmes

1.3 Research methodology and process

This research has been implemented by two complementary methods of data collection. Firstly, by reading and studying the relevant documents, reports of the agencies involved and secondly, by interviewing key contact person(s) of each organisation. Reading the documents prior to the interviews helped the consultant to enrich and finetune the interview questions. The topics of these interviews are presented in **annex 1**. In some instances the researcher sent clarifying questions to the interviewee after the interview for extra verbal information and, at times, the sharing of additional documents for the mapping.

The selected donor agencies were members of the SDG2 roadmap gathering. All these agencies have been approached for their inputs and participation. Not all of them participated for unknown reasons. On the suggestion of the secretariat of the donor platform some other countries were approached gauging their interest in this mapping exercise to which some of them responded positively. An overview of agencies that participated is shown in **annex 2**.

A limitation of this inventory is that not all result reports were available for this research. It is also quite possible that more reports are available and that therefore this inventory has been based on an incomplete set of documents and reports. It is however expected that on basis of the available reports a sufficiently reliable and informative overview has been arrived at that can be used for further steps to better results reporting on SDG2 and be used for further work in the SDG2 roadmap group.

1.4 Voluntary National Reviews

To date more than 100 nations have already conducted Voluntary National Reviews (VNR) in their countries, tackling the SDG status at country level. To the extent that they were directly or indirectly referred to in this donor mapping exercise, it was clear that they are fully separated from the usual, conventional reporting by the responsible agency for international cooperation on the results of their projects and programmes in developing countries. VNR have a more popular character and content, they have a more narrative style, focusing on activities and outputs, and spell out policy directions and long term goals.

Conducting these VNR has taken a lot of resources, time and energy, because of the involvement of many and very diverse national and local agencies, including civil society and private sector. The useful lesson from these VNR for this donor mapping is the demonstration of the dependency on these diverse implementing partners, which often have little or weakly developed M& E and reporting capacities and experiences. This seriously restricts the quality of data collection, analysis and reporting at project and aggregate level. This has its parallel with the implementing partners in development projects, which also struggle with their M&E skills and reporting products. Apart from this observation VNRs have not played a further role in this donor mapping exercise.

2. Key findings

Hereunder the main findings will be presented. In general, all agencies studied and persons interviewed recognised the need to improve the quality of results reporting. They also observed and agreed that improvement of this reporting takes time, being part of a slow but steady process of enhancing the quality of planning and M&E. It is work in progress where different agencies vary in pace and intensity. There are no conflicting views on the need for enhancing this reporting quality; the agency history of results reporting varies; political debates in parliament and society affect the reporting focus, orientation and dynamics.

2.1 Driver of results reporting

Four drivers have been distinguished and discussed with the different agencies: **accountability; communication; learning and steering.** Accountability and communication have a more external function and dimension. The reporting concerns the transparent demonstration of results achieved and spending of funds to the parliament, civil society organizations, the general public and tax payer. It generates inputs for debates on the relevance and value for money of the development budget. Learning and steering have a more internal functions, geared towards constantly improving the effectiveness of development interventions; producing a better understanding of what works or does not work, and why (not); and how best to spend funds in order to generate more effective results and changes in the situation of target groups. The focus of learning is on outcomes and impact level results.

This is the situation for all agencies involved in this mapping. Except for the Gates and Rockefeller foundation that use private money and who are not legally obliged to report to the parliament. Their focus is clearly on the learning and steering; learning feeds into adaptive management. Both organizations refer to accountability as a driver for learning and constant improvement.

It was repeatedly stated that these different drivers for reporting are not conflicting with each other; they may well act complementary to each other; co-existing and mutually reinforcing. But they generally lead to different types of reports about results achieved, using variable mixes of **qualitative and quantitative** information.

✓ **Qualitative** information is more used for accountability and communication purposes for the different audiences. These groups or stakeholders better appreciate and prefer good, reasonably validated stories and narratives, which easily feed into debates about development aid. The interviewees remarked that generally such audiences often have difficulties in appreciating **quantitative** results achieved. They poorly connect the results of their own agency – how many persons reached, who now have less hunger - with the global figures of, for example, FAO on persistent numbers of hungry persons in the world. So, whereas part of these external audiences may actually like quantitative and impressive numbers, the causal linkages with own development programs is too often poorly made and understood. Many agencies therefore have a preference for and greater reliance on credible qualitative information, using a

choice of case stories as illustrations of results, because these feed better into meaningful debates on the quality and effectiveness of development interventions and investments.

- ✓ **Quantitative** information serves better for internal learning and comparative analysis in order to arrive at proper decisions on which interventions and funding yields most value in terms of actual results realized. Properly measuring and counting the numbers of persons that have improved their food security related situation is the only means to agree on the effectiveness of different (types of) interventions. Though agencies also agree that they still face many challenges with regard to properly collect this quantitative information, they nevertheless acknowledge its relevance for learning. A simple qualitative story without numbers is not enough.

2.2 Need for evidence based information

All agencies agree to a more or lesser extent that the quality of result information must be based on credible and plausible evidence. There is an ongoing effort for more robust and rigorous evaluation methods and approaches in order to feed into better learning and adaptive management processes. The expectation and focus is not on scientifically sound and statistically grounded information, but they state that results must be better based on underlying surveys, studies and/or samples, which make information generated verifiable. The data collected and analysed must be sufficient to allow for **credible and plausible** findings and conclusions on effectiveness of development interventions with a relative modest set of data and indicators.

At the same time a number agencies (USAID, Gates Foundation, Italy, a.o.) currently undertake efforts to support national statistics bureaus in quite a number of countries in collecting relevant information in different sectors in order to understand what is happening. This effort must be seen as a parallel process to collect and generate evidence based information for learning. Better national statistics may function well to create a data base that describe national context and potentially be used for national policy decision processes, but it is not relevant for current learning and reporting, the topic of this inventory. Improved performance of national statistics bureaus may in due time form a highly valuable secondary information source, but program related learning at present needs a separate M&E system or practice that will also feed into aggregate reporting and comparative analysis.

2.3 Reporting as the final step of the Planning and M&E process

Reporting cannot be seen in isolation. It is closely linked with proper planning, smartly formulated objectives and targets, well selected and agreed upon indicators at all different result levels and a properly designed and implemented M&E process, in terms of skills, capacities and resources. The quality of reporting is strongly linked with the quality of the indicators and the M&E system, and the clarity of the stated objectives, targets and achievable, realistic and smart indicators.

All agencies make a strong reference to the need for proper planning and design of programmes; the presence of strong intervention logics, either described in a Theory of Change, impact pathways or Logical Framework. In that light also frequent reference is made to the need to arrive at a better consensus of what constitutes reasonably proven and effective interventions (see 2.7).

2.4 Dependence on weak M&E system and practices

All funding agencies depend on the M&E skills and capacities of their implementing partners on the ground as the basis for data collection, analysis and reporting. Across the board the quality of the existing M&E systems and practices leave much to desire: little time is spent for M&E; the M&E function is often poorly resourced; staff capacities are limited; lack of institutional recognition for proper M&E as a management tool; collected information focuses more on outputs than on outcomes or impacts; often the emphasis is on qualitative than on quantitative information. Therefore the quality of result level information is weak, very diverse and not sufficient to produce relevant, credible or reasonably validated information on results. Consequently, the quality of reporting suffers as a result.

The exceptions are the Rockefeller Foundation and the Gates Foundation who have partnered with research institutes that support them with the monitoring of results in the field. That support leads to more harmonized M&E and a much better quality reporting.

The implementing partners are assisted in various ways by the agency HQ, assisted by country teams and embassies, and often specially recruited M&E experts. But at present the M&E skills and capacities are still weakly developed. Agencies recognise the need to invest in better M&E practices on the ground, but across the board the concrete support of senior management to make significant and quick progress is often lacking. M&E strengthening is generally recognised as a process which takes much time and many small steps.

2.5 Methodological notes

A small number of countries have elaborated methodological notes that guide the M&E process for results measurement. These notes developed at HQ level describe range of selected indicators, the underlying rationale of each indicator; technical content and working definition; the expected disaggregation; relevant data sources; points of attention to avoid double counting; impact pathways or intervention logics; specific examples for illustration purposes; possible overlap with indicators used by other agencies (latter applies only for EU), etc. These notes serve to better align M&E of implementing partners, to facilitate aggregation and offer a more reliable scope for comparative analysis and learning in each agency.

The level of detail of these notes varies, but they generally reflect

1) a highly crucial process of generating sufficient internal consensus and common understanding at HQ level concerning these indicators. The underlying debates are usually intensive, as they show many different angles and interpretations. The key question is: What is included in selected indicators and what not?
2) a support for implementing partners in the field concerning which information to collect and how to collect for the commonly agreed upon indicators. The use of these notes is generally seen as a delicate balancing act vis-à-vis the existing M&E practices of the implementing partners. Usually, these notes are not seen as prescriptive, but they need to be gradually built and merged into the local M&E systems. These are quality investments that take years to produce the required results.

Sharing and agreeing on these different notes among the funding agencies and exchanging experiences with regard to its application on the ground may well serve for alignment and harmonization among them, and provide mutual support if and where needed.

2.5 Use of external evaluations

External evaluations serve as an additional tool to produce reliable and validated information, but its use is still limited for aggregated reporting. Broadly spoken two different types of external evaluation can be distinguished.

1) **On a program scale**: Mid-Term Reviews that directly feed into the current implementation phase of projects. They serve well the learning purpose by also bringing in an external view. Final evaluations have a lower connection with ongoing information, unless they are actively used for improved program formulation in a next phase. Various efforts to improve the quality of external evaluations; better TOR; better linkages with DAC evaluation criteria, more time, etc.

2) **On a global or sector scale:** looking at the aggregated results of various external evaluations or examining the sector wide lessons learnt by synthesizing the findings of (a sample of) all FS related programs in target countries. That level of evaluations and learning often takes place with the use of specialized institutions – internal or external – and they generate lessons learnt on a more aggregate level, feeding into the better design of effective, future programs and intervention logics. DGIS – IOB; RF and RTI (Research Triangle Institute), Gates and Brooking Institute, USAID.

Both types of evaluations are complementary and should be seen as 2 separate mechanisms to improve the effectiveness and quality of Food Security programs at different levels.

2.6 Consensus on terminology

The terminology as a basis for proper results reporting varies, it is at times vague or it may easily lead to misinterpretations. As a result this hinders the possibility and/or quality of aggregation. There are a number of dimensions of terminology that have been observed during this mapping and that need further consideration.

The key examples are:

✓ **Result hierarchy or levels**, distinguishing **outputs, outcomes and impacts**. As such there is a reasonable and sufficient recognition and agreement on these 3 key levels; there is also a fair consensus that for result reporting the focus should be mostly and preferably on **outcomes** achieved. But the way in which outputs and outcomes are described is not consistent, using different words or terms, which may be differently interpreted by different readers and agencies. Most agencies consider **impacts** as relevant for result reporting by implementing partners, as they are closely associated with long term results that can only be measured and reported on towards the end of the implementation period. Impact is rather the subject of external final evaluations. Interestingly, Rockefeller Foundation made more explicit attention to its contribution to impacts in connection with longer investment cycles, 5-10 years.

- ✓ **Differentiating outcomes**: within this outcome category it is useful to make a distinction between different stages or steps of outcomes, which can be positioned along the time-axis of often long change processes. In principle these change processes should also have been described in its Theory of Change or intervention logic of the program. Generally, references are then often made to descriptions like short and long term outcomes, or to intermediate outcomes. Whereas, the short term outcomes describe the first modest but necessary changes needed to arrive at, the content of the longer-term outcomes are more transformational. These latter outcomes have a better linkage with SDG2 indicators and should then be described with a clear picture of the final situation that it is aimed to arrived at. Some agencies already formulated such final pictures.

- ✓ **Annual reports** usually describe activities, interventions and outputs. On a limited scale outcomes form integral part of such annual reports, which enrich the quality of reporting, learning and communication. It would be relevant if more agencies would join this effort to include outcomes in their annual reports. The agencies that already started to include outcomes in their annual reports in a more systematic way are DFID, DGIS and USAID.
  - o **DGIS** is making concrete steps towards including outcomes of previous year in its annual reports
  - o **DFID** includes outcome data in the current annual reports, looking back 2-3 years;
  - o **USAID** produces a separate snapshot report, updated every next year, on outcomes trends over a longer period since start of Feed the Future program, 2011.
  - o A particular problem experienced with regard to outcome reporting is the timely availability of this information. Whereas output information is easy to gather and to update, this is less so with regard to outcome information. Monitoring and Evaluation of outcomes takes more time or requires a separate measurement effort by implementing partners (or external evaluators) so these results may not be available before the annual reporting deadline.
- ✓ **Consistent result terminology**: In spite of adherence to **DAC/OECD glossary** on these three result terms, the precise description of these result levels varies, using terms like reach, access, use, support to, enabling, which does not make it clear to which result levels (notably the distinction of output and outcomes) this refers to. A term like 'reach of a project' may be connected to all three result levels, whereas the DAC/OECD glossary clearly limits this term to outputs only. More agreement is needed. This also refers to a more important dimension of reaching consensus on the content of effective, proven interventions in different sectors relevant for SDG2, see hereunder 2.7.

- ✓ **Proven versus 'experimental' interventions**. An important dimension in this regard is the distinction between two different types of interventions. For the **nutrition sector** already more consensus exists on those interventions or activities that will automatically lead to the desired result: such as for example nutrition support to children under 5 to combat malnutrition. In such a case the number of children supported, although technically an output only, nevertheless equates the practical meaning of an outcome. This nutrition intervention is a proven activity

that logically leads to the intended outcome, less malnutrition. The situation is often different for the **agriculture sector**. The choice of interventions or activities (as part of the TOC or LF) will not automatically lead to or guarantee better performing and more food secure farmers, because the program has a more experimental character where learning still is prominent and needed. In such a case the simple output statement as the number of farmers trained is not equal to the outcome as the number of farmers who then apply new knowledge and/or practices on their fields. This distinction between proven and experimental programs must be considered in the proper appreciation of the result level terminology.

2.7 Consensus on effective, proven interventions

An important point of reflection, frequently mentioned during the interviews, was the need to arrive at a greater consensus of what actually constitute effective, reasonably proven interventions, having impact on the ground, based on the long yearly experiences of the various agencies across many countries, programs and contexts. The point is that if more consensus would exist then the measurement of results could be more harmonized too, using similar indicators and methodologies, making aggregation easier to arrive at.

The Rockefeller Foundation mentioned a particular interesting aspect of developing this consensus in relation to the choice of **farmers**. Talking internally about the challenges in the food security sector, the question was raised which (smallholder) farmers are the best target group to contribute solving food security problems. Should it be about the relatively bigger or smaller farmers? Which size of farm or type of farmers should preferably be targeted to get the best results? No internal consensus exists as to that key question. And apparently, as of now, no evidence is available to answer that question.

The Gates foundation in its Executive Summary of **Ending Rural Hunger** initiative "*Mapping Needs and Actions for food and nutrition security*" defines four key areas for collective action for real impact on FNS needs in multiple countries:

1) Integrating sub national, national and global food and agricultural commodity markets
2) Achieving agricultural intensification that is environmentally sustainable and resilient
3) Delivering new advances in location and crop specific research, technology and extension services
4) Transforming family farms from subsistence enterprises to small scale commercial businesses.

These four areas may serve as an overall guide for developing this needed consensus. Interestingly, the first area has a dual dimension as it highlights the need to look at developing and developed countries at the same time and aim at reducing the distortions in domestic agricultural and trade policies.

2.8 Overview of key aspects of M&E and reports

Hereunder, a short overview is presented per funding agency that has been involved in this inventory. The focus is on the indicators used, the result levels and the presence of methodological notes that witness the debate and consensus emerging with regard to proper data collection and analysis of the selected indicators. This is all work in progress and improvements are made on a regular, yearly basis.

Even though care must be applied in the external appreciation during this inventory, it can be noted that agencies like DFID, DGIS, EU, and USAID are ahead with regard to their internal debate, consensus development, the efforts to improve M&E systems and the results reporting quality. FINNAID and GIZ are making progress with regard to the development and elaboration of methodological notes, but as yet no annual reports are available with outcome results focus. The other countries agencies are lagging behind, but are highly interested in the results achieved by other member agencies of the SDG2 roadmap group. Rockefeller and Gates foundation operate in a different way, partnering with specialized research institutes, rendering support in monitoring and reporting.

Agencies	indicators used	target group differentiation	focus on outputs, outcomes	use of external evaluations	Aggregation Yes/No	methodological notes made
Australia	Employment rates; more DCED oriented	Gender;	outputs		No	No info
Austria	Contributed to third EU biennial report, but not yet published	Gender, age, disability ; location specific	outputs			No info
DFID	# of persons achieved FS # of persons reached with nutrition programs land rights improved for # persons # of farmers with increased yields # of farmers with increased incomes	gender: women children under 5, breastfeeding and pregnant women no	outputs, outcomes, as from 2017-18 outcome trend over last 3 years 2015-18		Yes	no methodological notes that promote alignment, though external expert inputs are provided to build more robust M&E practices..
DGIS	# of malnourished persons reached and improved food intake # of family farms that doubled their production and/or income # of farms with improved access to input and/or output markets # ha of land eco-efficiently used; part of improved watersheds; more resilient to shocks # ha of land have secure rights # of improvements in key (inter) national FNS policies and laws; # idem policies implementation # of major systemic linkages in FNS realized or improved	gender: women	outputs, outcomes	Yes, IOB food security evaluation	Yes	yes, very detailed methodological notes, plus results and indicator framework
EU commission	# of persons benefitted from nutrition programs # number of persons have secure land tenure # of persons received advisory services # ha of land have improved, sustainable management practices	women and children gender	outputs, outcomes		Yes	Yes, fairly detailed notes. Plus document on Intervention logic and indicators

	# of persons with access to markets # of food insecure persons received assistance	Gender; age categories, location				
Finland	# of persons with improved food security # of smallholders with sustainably increased production and climate smart agriculture # of smallholders with secure access to land	Gender	outputs outcomes	no direct use for results reporting	yes	yes, fairly detailed notes on indicators, incl TOC obs: outputs actually describe outcomes. Details of data collection in the field still to be developed.
France	# of family farms benefited from projects in SSA (EU report) # of farmers with increased yields # of farmers with increased incomes # of smallholders with sustainably increased production and climate smart agriculture # of smallholders with secure access to land	Gender, youth	no info	no info	no info	no info
Gates Foundation	# children under 5 with wasting and stunting HDDS and HFIAS (and other FAO indicators) Cereal yields per ha % of land dedicated to improved varieties	gender, age	outputs; outcomes & impact	external program evaluations	no info	a **tool kit** with 80 highly diverse indicators of the ERH program on 3 FS dimensions: needs, policies and resources; supported by Brookings Institute
Germany	# of persons with reduced hunger and malnutrition # ha of land have improved, sustainable management practices IDDS and HFIAS Policy change FS related	gender direct, indirect and non-beneficiaries	outputs outcomes			yes, fairly detailed; clear ref to **adoption rate**; incl impact pathways
Italy	# ha of land covered with crop # farmers with higher income # weight of crop sold		outputs	yes	No	Busy developing guidelines for result based approach
Rockefeller	Great number of diverse Value	gender, youth	outcomes	N.a.	yes	M&E and learning process

Foundation	Chain indicators, engaging with all relevant VC actors.		impact contri-bution			outsourced to RTI
USAID	# of households that are food secure # of children under 5 reached by nutrition programs and/or free from stunting # of health facilities with established capacities to manage under-nutrition # of farmers who applied improved technologies or management # of ha under improved land management practices # of farmers with increased yields total value of smallholder sales or earnings in market	gender; includes use of women's empowerment in agriculture index	outputs outcomes outcomes trends across FtF program; presented in snapshot reports. 2 reports (2017 and 2018) cover **6-7 years** trends	FtF global performance evaluation	yes; global and country wise	highly detailed FtF Indicator Handbook with reference sheets per indicator. Different categories of indicators: performance, Zone-of-influence; national level; total of 54 indicators; plus 25 context indicators..
Switserland	# farmers who increased production (crops, livestock) # of persons with secure access to land # of producers who apply water efficient practices on # ha # different policies related to FS, nutrition and access to resources # persons with diverse diets	Gender; age groups	outputs outcomes	no info	no info	Mandatory ARI: Aggregate Reference Indicators, complemented with Indicator Sheets (sort of meth note)

Comments to this overview

General

- Except for Gates Foundation all result information is linked with programs funded by the national development agencies and the results are therefore attributable on a reasonable scale to these programs.
- Indicators mentioned are not necessarily used consistently; agencies change and adapt indicators over time based on lessons learnt and different intervention approaches developed and applied.

Methodological notes: only stated as yes, if these notes were made available during this inventory and were actually discussed

Finland: HQ staff did an extensive pilot to collect and aggregate all result information of the 40 programmes in WASH and FS sector. That gave a good impression of time and labor involved. This led to the first annual, but unpublished report.

France: information only from VNR report 2016 & 2017 and EU 2016 results report. This inventory has used the description of the technical content of intended results to formulate the indicators that may be used for M&E and reporting

Gates Foundation implements the Ending Rural Hunger (ERH) project, measuring 80 indicators at national level in 3 FS related categories: FS needs, policies and resources. They use detailed indicators as defined by institutes like FAO, IFAD, WB, CGIAR, and many others. The indicators included are the ones that overlap most with those of other agencies

Italy: indicators mentioned were used by an external evaluation on Value Chains. Not directly applicable for country status as such.

Rockefeller Foundation has partnered with Research Triangle Institute to manage and facilitate an in-depth **learning** process, focusing on 3 VC in 3 African countries, on a quarterly basis, extending over period 2016-2018. The VC activities and results do not describe FS linkages.

USAID: # of farmers who applied improved technologies or management can be equated with the adoption rate of farmer; see reference made by Germany. The country based aggregation summarizes results of all different USG agencies funded programs. Indicators mentioned come closest to SDG2 and are most frequently used in reviewed aggregate reporting

Switserland: information on Aggregate Reference Indicators sheet only; no access to actual use in annual reports 2016, 2017, 2018

3. Research questions

With the previous key findings in mind, the stated research questions can now be answered.

3.1 Which result areas and indicators (at different result levels) do the funding agencies report on in relation to SDG2 targets and indicators?

SDG 2 indicators (see total number of them) are hardly used. The formal SDG indicators are rather seen as the responsibility of the national government and less of the international development agencies. SDG indicators are used in the VNR. Moreover, in many cases SDG indicators are seen as too global and less applicable for own programs. Agencies prefer their own indicators, tailored to the type of projects and programs they fund and implement, serving their learning and accountability purposes. There is overlap with some of the SDG2 indicators, but only to a limited degree.

The indicators mentioned (see overview) are those that are most clearly linked with SDG2. Several agencies also use indicators that belong to other SDGs, closely related or linked with SDG2. So the practical overlap is biggest with regard to the SDG 8, PSD interventions, dealing with market and value chain development and interventions; these programs directly contribute to income security, employment and job creation; functional markets; access to inputs and services. This all fits well in the access to food pillar of the Food Security concept.

Hereunder follow the result areas that by and large are used by most agencies and which could be seen as the possible starting point for reporting improvement and aggregation. For some of these indicators the result level can be clearly distinguished between outcome and impact attributable to specific programs. But in many cases the specification of these result levels depends on the content of the program. For example, the impact definition of a soil management program and a food security program vary, even though both programs may refer to improved food security as their ultimate goal or ambition. But the level of attribution (or contribution) of course varies greatly, the former only very indirectly and the latter fairly directly. As much as possible they have been ordered along the result hierarchy logic.

a. Outputs

Number of persons benefitting from nutrition programs. This result area concerns specific vulnerable target group categories with greater risks of malnutrition that need special attention to ensure the quality of food consumed: pregnant women and children under 5 are then mentioned. In many cases nutrition programs are based on more robust and already proven intervention logics, meaning that this indicator is an output indicator which will logically lead to better nutrition levels as an outcome.

Target group categories: direct and indirect. This refers to the distinction that agencies make between 1) persons they directly work with and consider as their direct target group and 2) persons who may also

benefit from to the program, but in a more indirect manner and are therefore considered as the indirect target group. Usually, the reporting is only on the results – outcomes - for the direct target group; the results for the indirect groups are not or under- reported. As a result this underreporting underestimates the total number of people who may have benefitted of the implemented program, which negatively influences the total result picture. Secondly, this may also illustrate a flaw in the intervention logic, which actually should already describe these intended change processes not limited to the direct target group only, but referring to copying or scaling processes. Such processes may as well have a systemic character, which thus far remains hidden or implicit in most cases. The indicator handbook of USAID and indicator sheets of GIZ may well be taken as illustrations of these different categories for which outcome measurement is relevant. So, this reach of target groups is also applicable for the higher result categories: outcomes and impact.

b. Short term outcomes

Adoption rate. This indicator describes the immediate or short term outcome change as the direct result of the implementation, confirming or not the direction of the intervention logic. Remarkably, only GIZ, Germany mentions the adoption rate as a critical indicator. The measurement of this indicator is fairly straightforward and, ultimately, it will be helpful to make a more credible estimation of number of persons reached or hectares of land improved.

Number of farmers with secure access to land. Improvement of land rights or tenure is seen as basis for improving farms, production and access to better incomes. Secure access to land is the necessary condition for farmers to start improving land or soil conditions, or invest in farm tools and equipment for raising yields and incomes. It is therefore considered as an outcome indicator.

Surface of land under sustainable management practices. It is evident that programs use a big variety of land (or soil) management practices that foster sustained yield levels; it concerns practices that promote soil fertility, water and soil conservation, crop rotation, agro-forestry; and a range of Good Agricultural Practices. The particular choice and combination of practices from the available list of practices have a highly contextual dimension. Agreement exists that such practices are needed for sustaining higher levels of production. This is also referred to as meaningful intensification. It can reasonably be stated that the proper choice and use of these practices are also favoring climate smart farming. The open question for measuring this indicator is still what level of quality of application of these practices is seen as sufficient to count a field as sustainably managed? In most cases this indicator is an outcome indicator, because it refers to the proper adoption of recommended practices by the farmers, needed in order to contribute to higher yields (= longer term outcome). This indicator corresponds best with SDG 2 indicator 2.4.1. "*the proportion of agricultural area under productive and sustainable agriculture*".

c. Long-term outcomes

Number of farmers who increased production. This indicator is strongly and narrowly connected with the 3 previous ones. Yields raised will directly contribute to availability of food and access to food. In the intervention logic it is the longer term outcome as a result of the above shorter term results. Only DGIS made reference to the level of intended yield increase, 100%. This corresponds with the **SDG indicator 2.3.1** "*volume of production per labor unit of farming/pastoral/forestry enterprise size*", stating a doubling of this productivity per 2030.

Number of persons that have increased their incomes. In principle this implies that smallholders have successful access to markets, selling their surplus, marketable crops (without differentiation to type of markets: local, regional, national, international) and thus increasing their income position. Better incomes then allow farmers to buy needed food they do not grow themselves. A better and more stable access to markets stimulates farmers to invest in the quality of land management (see above). Achieving incomes security contributes to better food security. Income security can be considered as the (longer term) outcome indicator contributing to food security impact. This indicator corresponds with **SDG2 indicator 2.3.2** "*Average income of small-scale food producers*", stating a doubling of this income position in 2030.

Number of persons with improved food security. This indicator is usually read and seen as the impact indicator, because it is at the end of the impact pathway to which various programs contribute. A number of agencies make use of the **HDDS and HFIAS** type of FS indicators**,** developed by FAO. These are globally accepted Food Security indicators, but agencies do sometimes adapt these to their organization specific interests or views. Germany uses I(individual)DDS instead of H(household)DDS. DFID mentioned the feasibility of use of **FIES** (Food Insecurity Experience Scale (a variation of HFIAS), which matches the **SDG2 indicator 2.1.2** "*prevalence of food insecurity*". These indicators describe results at higher impact level of FS programs. They have been widely validated, could well be connected with SDG nutrition indicators and are relatively low cost to collect.

For more agricultural programs these FS indicators are seen as sound proxy indicators for measuring the final impact of programs. In that case the attribution dimension is more complex and tricky. Their use must be combined with the other chosen specific (short & long term) outcome indicators of such programs.

Other discussion points

In addition, some other indicators have been discussed during this mapping exercise, including the target group differentiation. They may serve the purpose of measuring the results in a more specific way and stimulate learning and steering the programs.

- ✓ **Cost-effectiveness**. Although the need to look more at cost-effectiveness is underwritten, no systematic efforts are made as yet to use this as an indicator. Translating program budgets for agricultural development into an indicator of x euro per farmer is seen as potentially relevant, though with hesitations. Two statements with regard to its usefulness emerged from this discussion. Firstly, it may serve a learning and management purpose where sufficient consensus has emerged of what constitute effective programs or interventions; this is very sector

dependent. Secondly, it may serve for more informative comparative analysis and learning in those programs that are implemented in roughly similar contexts. At the moment it is not regarded as a proper standard for judging on the effectiveness of a program at its start of implementation. Such a standard may however better apply for more routine operations, where established best practices have already emerged, such as for example in nutrition and WASH related activities. For agriculture its use may as yet be limited.

✓ **Food Security related policy change**. During this mapping very few examples of reporting were found in which reference was made to results with regard to policy changes in Food Security sector. Policy change may figure in Theory of Change or Logframe. And in results frameworks and indicator sheets of DGIS and USAID reference is made to indicators for policy change, but no practical reporting examples have been found. The only policy related indicator with direct program reference was encountered in methodological notes of GIZ. In their case the policy change was connected with directly measurable change at the level of the target group, such as level of investments, yield improvements; access to services and inputs, and the like. In that way a much more reliable estimation of how many persons, farmers reached can be made; reducing the tempting tendency to overstate the reach of a policy change beyond the direct program.

✓ **Target group differentiation**. Where needed and relevant agencies do make and agree upon the need to differentiate their target groups. The rationale to look for differentiation when measuring and reporting on results is widely acknowledged, though actual practice of applying this differentiated monitoring and reporting is often more 'stubborn'.

- **Gender**: The practically universally applied differentiation concerns gender which applies for nutrition as well as agriculture programs. Almost all agencies report on gender as the percentage of women in the entire target group participating to diverse programs, so mainly output related. In terms of reporting on outcomes, this gender aspect is less well applied. For nutrition related programs differentiation is increased with age (children under 5) and specific categories of women (breastfeeding and pregnant).
- **Youth** is seen as a relevant category in agricultural programs in addition to gender. But thus far data about youth are not systematically gathered as part of M&E system.
- **Handicapped, ethnicity**, a.o. were also mentioned as relevant categories for differentiation, but no specific examples of reporting on these properties have been found.
- **Smallholders** are a relevant category to be included in reporting, but the term is more loosely used. There is no quantified definition (in terms of for example land cultivated or assets owned; number of livestock units owned), simply because such attributes vary largely, which is not surprisingly, given the highly diverse contextual situation in which farmers cultivate their farms. The only suggested quantified smallholder definition that would take into account this situational complexity is the specific **poverty profile** applied for that situation. The specific size of the farm would then matter less, but rather its relative position in the farming or market system.

3.2 To what extent are these different result areas and indicators comparable and aggregable? Looking at the aggregated picture at different result levels.

Not all donor agencies use the same combinations of these result indicators for their programs, which is logical in view of the different types of programs and objectives. In case of roughly similar programs, the precise formulation of the indicators used may still be different, but during this inventory it was clear that more alignment and harmonization is not a real bottleneck, once the precise rationale and technical content would be agreed upon. Two examples may serve for illustration purposes.

Firstly, **sustainable land management** is a broad umbrella term, consisting of the combination of many diverse practices (see list of such practices, shown by some agencies). It is clear that each program may choose and contextualize the best practices or techniques from that list and link its indicator to that choice. But if there is consensus that in all cases the final outcome of all these programs is sustainable land management by the farmer regardless the specific combination of practices, and that also agreement exist on the quality of application, then the hectares or surfaces covered as a result of program implementation is what matters. These hectares covered as the outcome can then be aggregated across all different programs and agencies.

Secondly, **smallholders** as a broad category of farmers, which also covers a wide range of properties or characteristics of farmers in terms of their farm sizes, type of crops grown, assets owned or access to key inputs and markets. Whether a farmer is considered a smallholder is indeed a key question, but it is evident that there is no single technical, narrow definition. Agreement exist that it concerns a relative concept, specific for a country or context of agricultural development. A potentially relevant attribute mentioned by USAID was the poverty profile of the farmer, which may be helpful for this smallholder classification. But if such basic agreement exists and is well considered in program design, then the results in terms of number of smallholders who will have improved their income and/or food security situation can be compared and aggregated.

In view of the highly variable M&E capacities and experiences of the different agencies and their implementing partners, it is evident that not all agencies will be able to handle the same number of indicators in an appropriate manner and with sufficient quality. A step by step approach should also be used in this process towards more meaningful and complete aggregation by all SDG2 roadmap members.

3.3 How could donor agencies improve their planning and reporting on results and contributions to SDG2 targets in a more meaningful way? These contributions must be plausibly attributable with implemented projects and/or programs

Consensus on effective interventions: Agree on what constitutes effective and plausibly proven interventions also in the agricultural and food security sector in which at present consensus seem to

lack. There may however already be more tacit consensus on what as a matter of fact constitute effective interventions. This development of consensus is best done on basis of specific examples of activities that are already fairly widely implemented by many agencies. The same may also apply for nutrition related interventions, though it is generally observed that in this sector more consensus has already been agreed upon and confirmed. The Ending Rural Hunger project of the Gates Foundation that defined 4 areas of common action (see 2.7), may possibly serve as an entry point for this discussion.

Consensus on robust evidence: this content based consensus should also include a common understanding on the necessary level of credibility, robustness of data collection tools and subsequent analysis, plausibility of evidence. The critical importance and contribution of more quantitative information to strengthen qualitative information should be part of this consensus.

Consensus on result terminology: Agree on the consensus with regard to result terminology with regard to the three basic result levels: outputs, outcomes and impacts. The use of precise and illustrative examples will serve as best tool to harmonize and align terminology, minimize confusion. Adoption rate is a sound entry point to feed into this discussion: jumping from output to (short term) outcome. It would also fit into a gradual process of improving M&E, emphasizing the development steps from short to longer term outcomes. The planning tools like TOC, LF and impact pathways should illustrate this consensus in result terminology.

Contribution and attribution: all agencies acknowledge that the measured results can never be attributed to their implemented programs only. Other actors and factors also contribute to these results. So, the true level of attribution should always be assessed as openly and transparently as possible in order to estimate the contribution they made. Sometimes the budgetary proportions of the different implementing agencies operating in the same sector & area can be used a proxy for such attribution estimation. It is also acknowledged that this estimation of the contribution is more complex at higher levels in the result hierarchy, i.e. at long-term outcomes and impact level. The quality and extent of details in the intervention logic or Theory of Change is very helpful to increase the quality of the contribution analysis. So, a better agreement on these intervention logics, in association with a consensus on the result hierarchy, will improve the quality of attribution. In more routine programs or sectors the level of attribution (or contribution) can be indicated more reliably.

Use of methodological notes: The methodological notes of the different agencies serve well for a critical debate and comparison in order to develop the common understanding of indicators, and the various associated steps of data collection and analysis.

Consensus on policy change results. This is probably a somewhat more unknown territory because few experiences may be available to build on. Monitoring of policy changes has always lagged behind. Nevertheless, the ERH project may again be used as a source of inspiration because policies are one of the pillars of this project; this refers to national & global policies. And, secondly, the GIZ reference to policy change measurement may deserve further attention because this refers to more practical illustrations of policy change on the ground.

Consensus on target groups and differentiation. This includes agreement on common understanding of direct and indirect target groups, so that credible estimation of total number of persons can be arrived at. Next, this may be relevant for a better estimation and measurement on the likely copying processes that take place for example at 1) target group level on the ground; 2) at value chain level through private sector; 3) at government level (local, district, etc) implementing relevant policies.

HDDS and HFIAS, or different variations of them; IDDS, FIES. These indicators have a well-tested methodology of use and application that could be copied by the various agencies. Agreement is needed in terms of the frequency used and how they will be merged with internal M&E systems of implementing partners. To facilitate this integration they should preferably not be used as stand-alone indicators, but in combination with program specific indicators; they can easily serve as a quick check at specific intervals in time (yearly) of the longer term impact results.

Use of the DCED measurement standard offers possibilities to harmonize results measurement. Though it belongs to SDG 8, its overlap with SDG2 is significant, certainly where it refers to number of farmers reached and income increases; and job creation (in value chains, potentially relevant for women and youth). It targets the food access in particular. DCED strongly advocates the proper and logical development of result chains and the steps needed to develop and strengthen the M&E systems and practices of implementing partners. This measurement standard is also useful for M&E of overlapping FS aspects.

Desired level of investments needed for M&E. In view of the widespread agreement or consensus on the need to improve the current M&E systems or practices of implementing partners, it would be worthwhile to exchange experiences and develop a common understanding of the appropriate level of quality of these M&E systems; their roles and responsibilities; what type of information to collect; which skills needed; what is the minimum level of resources – time, transport - needed; etc. Agreement on these issues would then lead to a more common understanding of the desired level of investments for better M&E systems. Much can possibly be learnt from agencies that already invested and progressed more in terms of M&E, and reporting skills. It would be relevant to exchange experiences how best to plan and execute better M&E, and what should be avoided (the do's and dont's).

Annex 1: overview of mapping topics

The interviews used a semi-structured approach with a number of topics and questions that could be contextualized if need arose. Based on reports received these topics could be adapted and/or expanded. The topics used were:

Key topic & question	Sub-questions
Main driver for result (measurement and) reporting	References made to accountability, communication, learning or managing.
Drive for evidence	Was the focus of M&E on Proving or Improving? Any reference to increased role of statistics for data gathering?
The result levels used	Which result level do you use or prefer: outputs, outcomes and impact?
Use of SDG Indicators	Which SDG2 indicators do you use? Or related indicators?
Type of result information	Which information - qualitative or quantitative – do you use and prefer, and why? What is the role of case studies in your reporting?
Target group differentiation	Which kind of differentiation do you make; refer to households, small farmers, gender, age; a.o.
Source of information	Is it about internal and/or external sources? Is there an internal evaluation unit? How do you use and integrate external evaluations in your reporting cycle?
Reporting focus	on own projects/programs only, or entire country?
Target audience	Who is the target audience of your reports? What kind of feedback from them on these reports?
Changes in reporting content and format	Any change in reporting over last years? Why?
Cost effectiveness	Any effort to link results with money spent or budgets? Value for money? Investments per farmer, hh, etc?
Project duration	What is the project duration used to measure outcomes and impacts in reports?
Other available documents	

Annex 2: overview of interviewed persons & agencies

Country & members	person	organisation
Australia	Tim Gill	DFAT
Austria	Waltraud Rabitsch	ADA
EU	Maria Paris Ketting	
Finland	Sanna-Liisa Taivalmaa Antti Rautavaara	FINAID
Gates Foundation	Ammad Bahalim	-
Germany	Anna Friedemann Elisabeth Mavrakis	BMZ GIZ
Italia	Marco Platzer Marinella Giannelli	AICS
Netherlands	Jeroen Rijniers, Paul van de Logt	DGIS
Rockefeller Foundation	Rafael Flor	-
Switserland	Ueli Mauderli	SDC
UK	Iris Krebber Karen Johnson	DFID
US	David Hegwood Jessica Cagley	USAID
SDG 2 Platform Secretariat	Laura Barrington	

Observation: with Belgium, Canada, France and Norway contacts have been established, but no interviews have been conducted.