Discerning the stars:
characterising the myriad of sustainability assessment methods

Wustenberghs, H.[1], Coteur, I.[1], Debruyne, L.[1] and Marchand, F.[1,2]

[1] *Institute for Agricultural and Fisheries Research (ILVO), Social Sciences Unit, Hilde.Wustenberghs@ilvo.vlaanderen.be*

[2] *Ecosystem Management Research Group and IMDO, University of Antwerp*

Keywords: integrated sustainability assessment, developers survey, characteristics analysis

Abstract

A myriad of sustainability assessment (SA) frameworks, metrics and tools have been developed. As the TempAg network aims to deliver resilient agricultural production systems, a.o. by comparing their sustainability performance, the first step was to identify currently used SAs and discern their characteristics. Therefore, from an SA inventory, integrated sustainability assessment (ISA) methods (assessing multiple dimensions) for agriculture were selected for an in-depth survey with the ISA methods' developers or users. A large variation in ISAs was found. Strictly reductionist representations were rare, but holistic ones ranged from less than ten to hundreds of indicators. Next to farm development, other (combinations of) purposes were found; a wide range of end-users; a spectrum of data collection, processing and scoring methods; and variate methods to combine indicators into an ISA. Stakeholder involvement in ISA development was found common practice, especially in the early phases, defining the sustainability framework and selecting the indicators. This first pilot activity shed some light on the complexity of ISA methods and the variability in their characteristics. Further research may reveal how they can be sufficiently enhanced to futureproof agricultural decision making.

1. Introduction

What is "sustainable agriculture"? How is it perceived in different regions and contexts? How can agriculture's sustainability be assessed? In trying to answer these questions, a myriad of frameworks, metrics and tools have been developed. Assessments originated top-down or bottom-up; with or without stakeholder involvement; aiming at farm development, food certification, policy evaluation, global reporting, etc. For TempAg, an international research consortium for sustainable agriculture in temperate regions (Gregory and Kougioumoutzi, 2016), one of the aims is to deliver resilient agricultural production systems, a.o. by comparing their sustainability performance. Therefore, the first step was getting grip on the currently used sustainability assessment frameworks, metrics and tools, how they originated and how different purposes resulted in different methods. This paper reports on TempAg's Pilot Activity 1.1.1, in which efforts for assessing sustainability in temperate (non-tropical) countries were surveyed.

2. Material and method

2.1. Inventory of sustainability frameworks, metrics & tools

In 2001 already, Riley noticed an "explosion" of indicators for agroecosystems, sustainable land management, biodiversity, social development, rural livelihoods, natural resources conservation, etc. Nowadays many of those indicators are used in more holistic frameworks, integrating several or all of the aspects mentioned. However, the universe of frameworks, metrics and tools for sustainability assessment is ever-expanding (Pope *et al*., 2013; Schindler, 2015). Any attempt at an assessment inventory of can therefore at best be comprehensive, but not exhaustive.

For this pilot activity we could elaborate on earlier compilations of frameworks, metrics and tools, such as the ones made for the SAFA framework (FAO, 2013), by the TempAg network and by ILVO (Marchand *et al*., 2014). These inventories were complemented through study of peer reviewed, grey and internet literature. The inventory currently contains about 170 sustainability frameworks, metrics and tools.

From this inventory, a selection was made for further evaluation. Frameworks, metrics and tools were included, if they were: (1) specific to agriculture or applicable with minor modifications; (2) developed in and/or applicable in temperate climates; (3) designed to assess sustainability. As sustainability is commonly seen to encompass at least three dimensions, economic, environmental and social sustainability (Brundtland, 1987; Schindler *et al*., 2015), *integrated* sustainability assessment (ISA) methods, assessing at least those three dimensions were prefered. ISA methods were not specifically selected on their scope, level or scale, although emphasis was put somewhat more on farm level assessments. The selection revealed 51 ISA methods, from all over the world, with broad ranges of scopes, assessment levels and data gathering scales, which were subsequently surveyed.

2.2. Survey of assessment system characteristics

How does one navigate between the myriad of sustainability assessments? How can one find the way to the right tool for one's purpose? Are there any dots and lines to make up a map? In other words: What are the key characteristics to describe frameworks, metrics and tools that may facilitate choice? The review of characteristics proposed to discern ISA methods, is discussed by Coteur *et al*. (2016). They selected 25 essential characteristics, which provided the basis for a survey on (1) the general ISA characteristics, (2) stakeholder participation in ISA development and (3) the use of indicators in ISAs.

Qualtrics Research Suite was used to build a web-based questionnaire. E-mails were sent out to the ISA's developers or users, inviting them to take part and providing them with a questionnaire link. Information on 38 ISAs was retrieved, i.e. a 75 % response rate. We feel confident that this sample is representative for the selected ISAs, that no specific ISA type or origin was left unsurveyed and that non-response was sufficiently random.

The survey responses were first analysed descriptively per characteristic. Second, relations between the characteristics were sought. For the continuous variables Pearson correlation coefficients were calculated. In the survey, however, most of the questions had multiple nominal categorical answering possibilities. These categories were converted to dichotomous variables (an option is used yes/no). Associations between these variables were determined by calculating tetrachoric correlation coefficients in SAS 9.4 software. The tetrachoric correlation - that rests on the assumption of underlying normally distributed variables (Bonnet and Price, 2005) - was preferred to the phi-coefficient - the linear correlation between underlying inherent dichotomous distributions (Chedzoy, 2006) - because of the calculation ease for many variables at once. Since Ekström (2011) ascertained a continuous bijection between both association measures, the underlying joint distribution should not have a substantial impact on the conclusions drawn from the analysis.

3. Results

3.1. Descriptive analysis of assessment characteristics

Of the 38 ISA methods, for which the survey was filled out, 20 were developed in western Europe and 14 at the international level. Only 3 originated from North and Central America and 1 from New Zealand. The distribution in the responses reflects the origins in the inventory, in which ISAs from eastern Europe, Asia, Africa and South America are scarce or even lacking.

3.1.1. General assessment characteristics

The survey results are discussed below per ISA characteristic.

- **Scope of the assessment:** 31 (80 %) of the surveyed frameworks, metrics and tools assess at least 3 sustainability dimension. Almost all methods assess the economic, environmental and social dimensions, 10 ISAs also assess the governance dimension. Other assessment dimensions mentioned include culture ("way of life"), plant cultural practices, animal welfare, entrepreneurship, innovations, multifunctionality and services.

- **Perspective on sustainability:** Only a minority of ISAs (7) looks at sustainability from a societal point of view, 16 take the farm's perspective. Most respondents ticking "other", indicate that their ISA takes mixed points of view, e.g. "both societal and farm", "farm and regional", "societal and distributer and farmer". Also the "value chain" perspective is mentioned.

- **Primary purpose of the assessment:** Reflecting our ISA selection criteria, farm development is the primary purpose (intended function) of almost 2/3rd of the ISAs. Research, reporting and communication are each mentioned for almost 1/3rd of the ISAs. For over half of the ISA's multiple purposes were reported.

- **Level of assessment:** Even more than farm development is a main primary purpose s, the farm is the main level of assessment (26 ISAs). Indeed, purposes such as identifying good practices, management optimisation or thinking and talking about sustainability are also supported by farm level methods. Field, chain, landscape and national/regional level are only mentioned for 8, 7, 5 and 5 ISAs respectively. For 27 ISAs (73 %) only one assessment level is reported.

- **Sector scope:** The majority of the ISAs are general, they can assess all farm types. Some are developed and/or mainly used in specific farm/production types, e.g. DEXiFruits, Ben & Jerry's Caring Dairy. Some ISAs consider more than farming, such as also forestry and fisheries (e.g. GlobalGAP, SAFA) or also processing of agricultural commodities (e.g. Field to Market).

- **System representation:** Only 2 respondents (5 %) claim that their ISA represents the agricultural system in a reductionist way, i.e. "few indicators are used to assess the sustainability of a whole system" (MESMIS and Sustainable Value Added). Half of the respondents (51 %) state a holistic representation, "reflecting the complexity of a system by using many divers indicators". 43 % state a combination of both, including the 3 ISAs that use only 8 indicators. As the economic dimension is handled in a more reductionist way than the environmental dimension, many ISA methods indeed use a "combination" of representations.

- **Applying users**: People carrying out the assessments are quite diverse. In 18 ISAs researchers are still involved in the implementation. Almost as important are farmers and extension workers (advisors, consultants). 17 respondents report that the assessment is a joint effort by several people with different functions e.g. farmer + advisor. Other applying users mentioned are NGO's or supply chain actors.

- **End-users**: Individual farmers are the end-uses of the result of 3/4^{rs} of the ISAs. The results of 1/2 of the ISAs can also be used in farmers' discussion groups. Only 3 respondents claim their ISA has a single type of user, for all other ISAs multiple end-users are foreseen (up to 8 types for the GRI G4 Sustainability Reporting Guidelines). Other end-users mentioned are students, policy makers, civil society, capital providers, operators in the supply chain, retailers, consumers, etc.

- **Time needed for data collection**: For only 5 ISA methods it takes less than 2 hours to collect the necessary data. For 14 ISAs data collection takes 2 to 4 hours (half a day), but for 12 ISAs it takes 2 days or more.

- **Data collection methods**: Interviews and self-assessments are both used in over half of the ISAs. Audits are used in 7 ISAs. Other methods include field measurements, animal welfare appraisal by vets, focus group discussions, surveys, public data, literature, etc. 17 ISAs use only one method, 20 use combinations of methods.

- **Indicator aggregation and weighting**: 2/3rd of the respondents indicate that the indicator scores in their ISA are aggregated (figure 1). Aggregation methods such as multi-criteria analysis, average scores per theme, simple sums and weighted sums are used. From the 22 ISAs that apply indicators aggregation, 15 weight the indicator scores before aggregation and again a variety of methods is described. A few methods leave the weights open, to be determined *ad hoc* by different users.

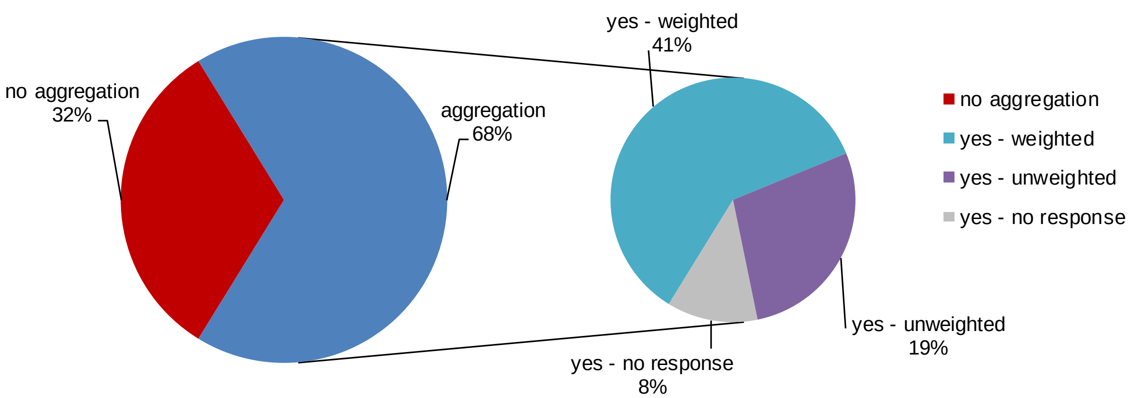

Figure 1: Aggregation of indicators scores and weighting in case of aggregation

- **Transparency**: Only 2 respondents state that no background documents are available about their ISA. Otherwise the ISA transparency seems quite well insured: for 10 ISAs documents are available on 5 topics, for 13 ISAs background documents are even available for all 6 topics mentioned in the survey.

- **Implementation**: "Is the assessment being implemented?" was answered by 34 respondents, of which 30 said "yes". 23 ISA's were implemented on project basis, of which 10 only on project basis, which might indicate that for 1/3rd of the ISAs implementation never went beyond the project were they were developed (yet). For the ISAs that are used by farmers, the respondents almost always indicate a combination with commercial use or certification use. Only 3 ISAs are implemented for farmers' private use only, outside a commercial/certification context. All 3 are linked to implementation on project basis.

3.1.2. Stakeholder involvement

Binder *et al*. (2010) defined 6 stages in ISA development and implementation:

1. *Preparatory phase:* defining context, goal and challenges (framework);
2. Indicator *selection:* choosing appropriate sustainability indicators, taking decisions on including interactions between indicators and weighting indicators;
3. *Indicator measurement*: quantifying indicators and processes (use of statistical data, surveys or categorized qualitative data);
4. *Aggregation of indicators:* deciding on whether or not to aggregate indicators, to which extent and how;
5. *Applicability* of the assessment results: getting the generated knowledge ready for utilization in practice;
6. *Follow-up:* reporting results, developing management advice, monitoring over time.

We surveyed stakeholder involvement in each of these phases.

Stakeholder involvement was revealed to be common practice in the first two phases: in 94 % of the ISAs, for which we received an answer, stakeholders played a part in defining the framework and in indicator selection. Stakeholder participation then falls back somewhat, but was in either phase still used in over 70 % of the ISA methods. Focus groups are the most frequently employed methodology for stakeholder participation.

In all phases researchers are the most frequently involved stakeholders (figure 2). In 2/3rd of the assessment methods, farmers were involved in the preparatory phase. Their involvement deceases as the development progresses, but reaches 2/3rd again, in the last 2 phases. Extension workers (advisors) mainly intervene in the 3rd and 6th phase, i.e. in indicator quantification and in follow-up/implementation. If involved, civil society (including NGOs) and policy makers mainly intervene in the early phases. Other stakeholders consulted are food chain and retail representatives.

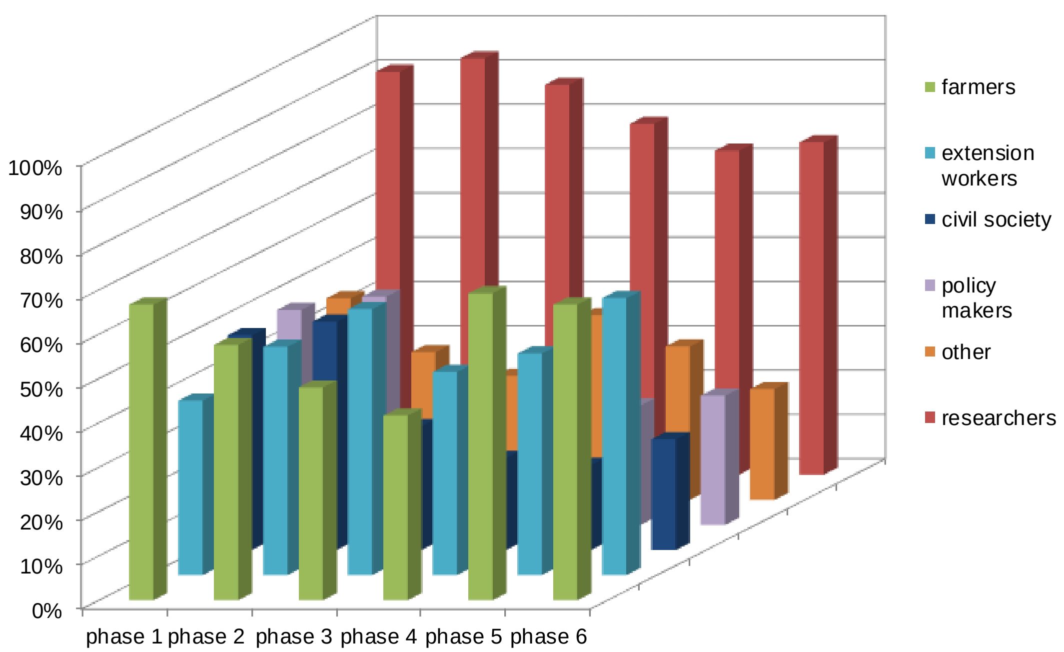

Figure 2: Percentage of ISAs in the survey in which different types of stakeholders are involved in each of the 6 phases of ISA development and implementation.

3.1.3. Indicator related information

33 out of 38 respondents answered "yes" to the question whether indicator related information is available (2 answered “no”, 3 did not respond). Only if this questions was answered affirmative, and respondents had stated before that a particular sustainability dimension is assessed in their ISA, they were shown the subsequent questions on the indicators in each dimension. The following analysis is thus based on a variable amount of responses (table 1 and 2).

- **Numbers of themes and indicators per sustainability dimension:** A large variation is reported in the numbers of themes used to describe each sustainability dimension, from only 1 to 25 (table 1). The number of indicators shows even more variation, from only 1 economic indicator in the OCIS Public Goods Tool, up to 300 social indicators in OXFAM’s Behind the Brands Scorecard. The smallest total numbers of indicators are used in the Fieldprint Calculator, the SAI Sustainability Performance Assessment, the TOA-MD 5.0 model (8 indicators each) and the Farm Route Planner (10 indicactors). On the other end of the spectrum, 700 indicators make up the Sustainability Monitoring and Assessment RouTine (150 economic, 200 environmental, 200 social and 150 governance indicators). The majority of ISAs uses more themes and more indicators to describe the environmental dimension than to describe the economic and social dimensions. The numbers of themes per dimension are quite well correlated and the number of indicators extremely well (table 4).

- **Indicator types:** For the economic and environmental dimensions mainly quantitative indicators are used, or a mix of quantitative and qualitative indicators. For the social dimension only few methods exclusively use quantitative indicators, for the governance dimension none do (table 2).

Table 1: Numbers of themes and indicators per dimension in the ISAs in the survey.

	Dimension	**Economic**	**Environ-mental**	**Social**	**Gover-nance**	**Total ISA**
	N responses	25	28	25	7	29
Themes	**median**	**4**	**6**	**3**	**5**	**15**
	min	1	3	2	1	5
	max	19	18	25	14	198
Indicators	**median**	**9**	**22,5**	**18**	**19**	**64**
	min	1	5	2	1	8
	max	150	200	300	150	700

Table 2: Indicator types used per sustainability dimension in the ISAs in the survey.

Dimension	**Economic**	**Environ-mental**	**Social**	**Gover-nance**
N responses	28	31	28	8
Primarily quantitative	50 %	52 %	14 %	0 %
Primarily qualitative	21 %	19 %	36 %	38 %
Equally quantitative and qualitative	29 %	29 %	50 %	63 %

- **Level of data input:** For all dimensions the farm and the farmer are the main levels of data input. The field, product or region levels are less prevalent in the surveyed ISAs. Other levels mentioned include the supply chain, community, a mix of levels for the environmental dimension and the farm family for the social dimension.

- **Data sources:** Farmers' knowledge is the data source most tapped in to by ISAs: in 75 % of the methods and for all sustainability dimensions. The accountancy is a source for economic data in 60 % of the methods, but also for environmental, social and governance data it is still used quite frequently. About half of the methods also need expert information. Field and site practices obviously are mainly used for economic and environmental indicators. Other data sources mentioned for the economic dimension are literature and modelling; for the environmental dimension expert systems; for the social dimension the community, regional sources, household survey, survey with farm workers; and for the governance dimension local policies.

- **Indicator scoring:** For the economic and environmental indicators, scoring systems based on benchmarks are most used (75 and 85 % respectively). Expert based monitoring becomes more important for the social and especially for the governance indicators. Several respondents report a mix of scoring systems within one dimension.

- **Reliability of data input and indicator validation:** Here non-response rates range from 18 % for the economic dimension to 37 % for governance. Do respondents feel this is sensitive information and thus are reluctant to answer? Or were "reliability" and "validation" insufficiently explained?
  The share of respondents stating that data input for all indicators is reliable is smallest for the social dimension. None of the respondents indicate that the data input for the economic indicators is doubtful. One does so for the environmental and 5 for the social indicators. Potential causes might be related to the data sources or the more qualitative nature of the social and governance indicators. If so, are these indicators less reliable *in se* or do the ISA developers/users feel less comfortable with qualitative indicators?
  About 2/3rd of the respondents state that the economic and environmental indicators in their ISA methods are validated. Only about 1/3rd does so for the social and governance indicators. Validation methods include resource data validated in previous studies, comparison with other methods, peer review, checking results with experts (e.g. accountants in case of the economic indicators) and participative group validation.

3.2. Relations between assessment characteristics

3.2.1. Correlations between numeric assessment characteristics

In the questionnaire a number of options were given for most general ISA characteristics. Many respondents ticked several options, indicating e.g. multiple primary purpose. The numbers of attributes per general assessment characteristic proved to be quite well correlated (table 3). The number of primary purposes (intended functions), number of dimensions considered, number of assessment levels (spatial scales), number of applying users (carrying out the assessment), number of end-users (using the assessment results), and number of ISA components for which background documents are available, all proved positively correlated. The correlations are not very strong, but many of them are statistically (very) significant. ISAs with more purposes thus usually also consider more dimensions, are assessed on more assessment levels, are applied by more users, can serve more end-users and have more types of background documents available.

Table 3: Correlations between the numbers of attributes of the general ISA characteristics

	N° primary purposes	**N° assessment levels**	**N° applying users**	**N° end users**	**N° types of background documents**	**N° phases with stakeholder involvement**	**N° stakeholder groups (median from 6 phases)**	**Implementation**
N° dimensions considered	0.407 0.012	0.475 0.003	0.366 0.026	0.480 0.003	0.257 0.125	0.148 0.384	-0.121 0.488	0.258 0.141
N° primary purposes	1	0.419 0.010	0.363 0.027	0.291 0.081	0.279 0.095	0.251 0.133	-0.020 0.911	0.355 0.040
N° assessment levels		1	0.303 0.068	0.442 0.006	0.303 0.068	0.115 0.498	0.168 0.335	-0.012 0.944
N° applying users			1	0.545 0.001	0.274 0.101	0.320 0.053	0.082 0.640	0.131 0.460
N° end users				1	0.427 0.008	0.465 0.004	0.137 0.433	-0.139 0.433
N° types back-ground docs					1	0.248 0.139	0.373 0.027	0.300 0.085
N° phases with stakeholders						1	0.126 0.471	0.102 0.565

Pearson Correlation Coefficients and Probability > |r| under H_0: Rho=0. Statistically significant correlations are highlighted in blue for probabilities ≤ 0.01, ≤ 0.05 and ≤ 0.10 respectively.

Table 4: Correlations between the numbers of themes (above the diagonal) and the numbers of indicators (below the diagonal) per sustainability dimension

Themes / Indicators	**Economic**	**Environmental**	**Social**	**Governance**
Economic		0.2737 0.2178	0.1851 0.4218	0.4409 0.3221
Environmental	0.6142 0.0018		0.9759 < 0.0001	0.7413 0.0566
Social	0.9014 <0.0001	0.8077 <0.0001		0.7750 0.0407
Governance	0.9889 <0.0001	0.9817 <0.0001	0.9962 <0.0001	

Pearson Correlation Coefficients and Probability > |r| under H_0: Rho=0. Statistically significant correlations are highlighted in blue for probabilities ≤ 0.01, ≤ 0.05 and ≤ 0.10 respectively.

3.2.2. ISA complexity

The correlations above point to some kind of continuum among the ISAs with increasing complexity. Marchand *et al*. (2014) proposed characteristics to discern the complexity of indicator based sustainability assessments at farm level. They observed a continuum between two extremes: the full sustainability assessment and the rapid sustainability assessment. Full SA tools make use of detailed farm data and/or expert information, need trained advisers and/or expert visits to gather the data, and are rather long and expensive in duration. Rapid SA tools represent the other side of the spectrum. They use farmer's knowledge or readily available data, allow an audit by the farmer or an adviser, and are relatively short in duration. Based on these observations, one might for our sample of ISAs expect a relation between the number of indicators in a sustainability dimension and the numbers of data sources, methods for data collection and levels of assessment. However, this relation could not be confirmed: no significant correlations were found between these numeric values.

Also between the total number of indicators in the assessment systems and the time needed for data collection, no relation was found (figure 3). Some combinations seem quite logical, e.g. > 2 days to collect the 300 indicators for OXFAM's Behind the Brands Scorecard. Some combinations seem counterintuitive, but can be explained by the method of data collection. For DEXiFruit, for example, < 2 hours suffice to collect the data to calculate 175 indicators, using existing databases complemented with expert knowledge. By contrast data collection for the TOA-MD 5.0 model takes > 2 days for 8 indicators, but the indicators need to be modelled. Not as much the number of different data collection methods seem to determine the duration of data collection, as the type of method.

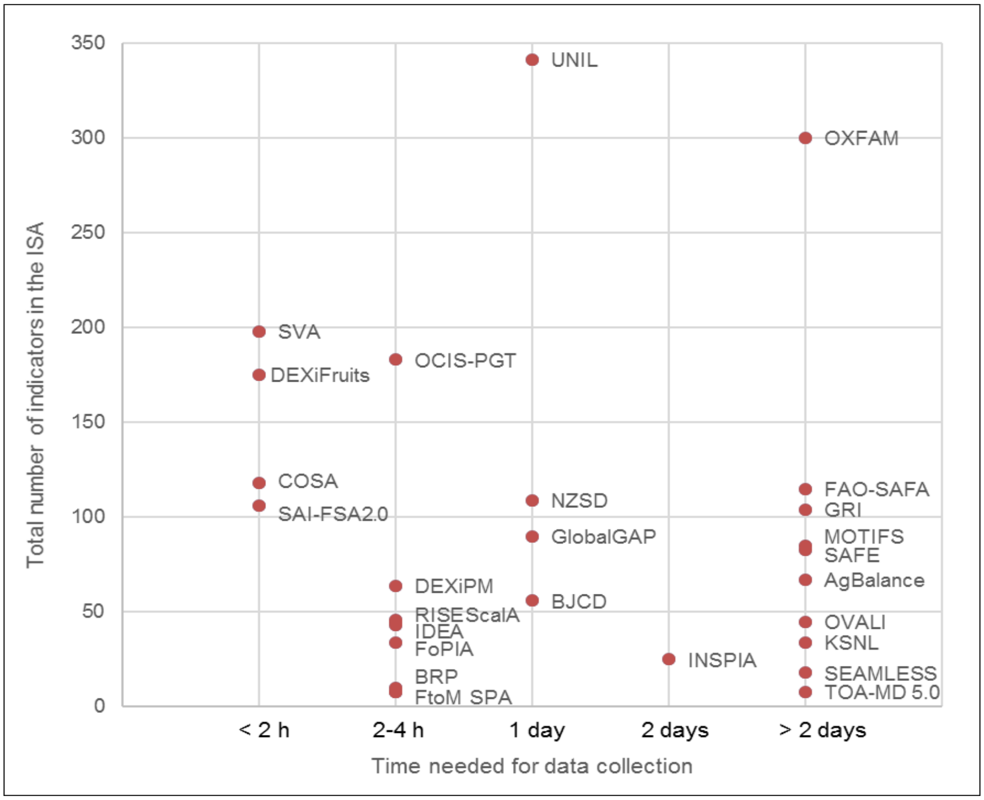

Figure 3: Time needed for data collection versus total number of indicators per ISA.

3.2.3. Associations with primary purpose and end-user

From the descriptive analysis it already became clear that not all ISAs cover all sustainability dimensions: some have a broader scope than others. The tetrachoric correlation analysis between the individual dichotomous ISA characteristics showed that the assessment scope is associated with both its primary purpose and its intended end-user. The primary purposes communication and farm development are strongly associated with the presence of an economic dimension (table 5). On the contrary, the certification purpose is associated with the absence of an economic dimension. Farm development is also strongly associated with having an environmental dimension, while the other purposes are not. Regardless of the end-user of the ISAs the environmental dimension is most prevalent (tetrachoric correlation > 0.98 for all types of end-users) (table 6). The economic dimension is most likely assessed if the end-users are policy makers, researchers or farmers in discussion groups. Individual farmers are not significantly associated with the economic dimension, probably because this dimension was significantly absent from certification systems and the individual farmer is an important end-user for those.

The social dimension is strongly associated with policy makers. The governance dimension is associated with research, policy makers and farmers in discussion groups.

Table 5: Associations of some general ISA characteristics with the primary purpose of the assessment

		Primary purpose									
		reporting		communi-cation		farm development		research		certification	
		Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²
Assessment scope: dimensions	economic	-0,068	0,834	0,976	0,041	0,669	0,018	0,269	0,423	-0,576	0,090
	environm.	0,977	0,227	0,977	0,254	0,999	0,052	-0,262	0,537	0,920	0,492
	social	-0,310	0,367	0,971	0,101	0,491	0,138	0,082	0,824	0,973	0,324
	governance	0,352	0,252	0,121	0,709	0,431	0,172	0,068	0,834	-0,964	0,220
Assessment level: spatial scale	field	0,160	0,593	-0,981	0,016	0,058	0,843	0,394	0,169	-0,967	0,149
	farm	0,061	0,831	0,238	0,419	0,481	0,064	-0,548	0,036	0,973	0,082
	industry	0,999	0,001	0,357	0,298	0,982	0,034	0,310	0,367	-0,973	0,324
	chain	0,689	0,010	0,307	0,312	0,501	0,096	0,487	0,090	-0,965	0,181
	nat./regional	-0,185	0,597	-0,139	0,696	-0,592	0,052	0,475	0,128	-0,966	0,267
	landscape	-0,976	0,050	-0,139	0,696	-0,592	0,052	0,727	0,011	-0,966	0,267
	other	0,352	0,252	-0,976	0,041	-0,158	0,609	0,068	0,834	-0,964	0,220

Tetrachoric Correlation Coefficients and Probability > Chi Square under H_0: Rho=0. Statistically significant correlations are highlighted in blue for probabilities ≤ 0.01, ≤ 0.05 and ≤ 0.10 respectively.

Table 6: Associations of some general survey characteristics with the end-user of the assessment

		End user: Who is using the results of the assessment?									
		individual farmer		farmer in discussion groups		extension workers		policy makers		research	
		Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²	Corre-lation	Pr > χ²
Assessment scope: dimensions	economic	0,179	0,585	0,571	0,055	0,125	0,692	0,987	0,008	0,604	0,039
	environm.	0,999	0,014	0,992	0,083	0,982	0,142	0,982	0,142	0,995	0,072
	social	0,012	0,974	0,020	0,954	-0,491	0,138	0,982	0,034	0,422	0,211
	governance	0,172	0,621	0,538	0,075	0,422	0,158	0,669	0,018	0,991	0,004
Assessment level: spatial scale	field	0,304	0,353	0,435	0,123	0,601	0,025	0,398	0,156	0,629	0,024
	farm	0,687	0,007	0,127	0,641	0,093	0,736	-0,108	0,693	-0,011	0,969
	industry	0,969	0,123	0,352	0,306	0,999	0,005	0,999	0,005	0,986	0,021
	chain	-0,092	0,774	0,102	0,733	0,289	0,325	0,520	0,065	0,571	0,050
	nat./regional	-0,570	0,064	-0,177	0,584	0,301	0,345	0,999	0,001	0,989	0,009
	landscape	-0,570	0,064	0,135	0,676	-0,009	0,979	0,592	0,052	0,421	0,194

Tetrachoric Correlation Coefficients and Probability > Chi Square under H_0: Rho=0. Statistically significant correlations are highlighted in blue for probabilities ≤ 0.01, ≤ 0.05 and ≤ 0.10 respectively.

The associations of primary purpose with assessment level point to different spatial scales being assessed for different purposes (table 5). The reporting purpose is strongly associated with industry-wide and chain level assessment, but the landscape level is absent. ISAs with a communication purpose do not use field level assessment (negative association). If the purpose is farm development, assessment can be performed at farm, industry or chain level, but not at landscape, regional or national level. The ISAs with a research purpose focus on landscape or chain level assessment, but not on the farm level. ISAs with a certification purpose, by contrast, are strongly associated with farm level assessment.

Concerning the end-users, the assessment level associated with individual farmers is the farm (table 6). This is probably linked with the certification tools in the survey that have the farm as assessment level. The larger spatial levels, landscape, or national/regional are not used for individual farmer's assessments. These level are rather associated with policy makers, who are also strongly associated with the industry wide level and with the chain level. They are not concerned with the farm or field assessment levels. Rather surprisingly, the extension worker (advisor) as end-user is strongly associated with the field and the whole industry assessment levels, not with the farm level.

Another interesting association is found between the individual farmer as end-user of the ISA results and the system representation. The association is negative for holistic ISAs (- 0.507), but positive for combinations between holistic and reductionist representations (0.646, both significant). This indicates that reductionism is important when farmers use ISAs. This is consistent with Schindler *et al*. (2015), who argue that reductionist methods might facilitate the communication of complex and complicated information, but also risk losing sight of the complex and often characteristic picture of reality and of what is important at the local level. Sustainability assessment should thus allow some complexity, but above all provide sufficient stakeholder interaction to understand the local context and to elaborate indicators that fully represent the analysed system, while remaining useful.

3.2.4. Stakeholder participation

Schindler *et al*. (2015) emphasize the importance of stakeholder involvement in ISA development. They recommend participation throughout all phases, from planning to final evaluation (see section 3.1.2). For the ISAs in our analysis, the number of phases with stakeholder involvement shows a significant positive correlation with the numbers of applying users and end-users (table 3). Stakeholder participation throughout the development process is thus linked with more types of users. The correlation evidently does not show the causality of this relation. Developing an ISA method that envisages multiple users, might require more stakeholder involvement or inversely, if stakeholders are involved in more phases, they might be more willing to implement the ISA, as suggested by Binder *et al*. (2010), Triste *et al*. (2014) and others.

A negative correlation was found between the number of phases involving stakeholders and the number of environmental and social themes (- 0.573 and - 0.559 respectively). This could indicate that more frequent stakeholder involvement might help to restrain the number of themes being assessed or maybe just to cluster indicators in a smaller numbers of themes. The number of indicators was not significantly correlated.

Moreover, one could imagine that stakeholders with different backgrounds involved in the early phases of ISA development, might result in more diverse purposes or themes taken into consideration. This assumption, however, is not confirmed:no significant correlations were found between de the numbers of stakeholder categories and either of the general ISA characteristics, nor with the numbers of themes/indicators. The only exception is a 0.60 (very significant) correlation between the number of stakeholder categories in phase 5 and the number of applying users. Also, the number of end-users, the number of assessment levels and the number of background documents all were correlated with stakeholder involvement in phase 5 (0.49, 0.35 and 0.43 respectively). This emphasizes the importance of diverse stakeholder participation in getting the ISA ready-for-use in practice.

The lack of association with stakeholder involvement was confirmed by the tetrachoric correlations with the individual ISA characteristic. This for instance showed that the intended end-users are not necessarily involved in the development. For ISA's used by individual farmers, farmer participation is only significantly positive in phase 5 (applicability). In phase 3 (indicator quantification) the association between the farmer as end-user and farmer participation is even strongly and significantly negative. By contrast, extension workers and policy makers are involved in most development phases of ISAs for which they are the end-users.

Finally, we checked whether stakeholder involvement improved transparency in the sense of the number of background documents available. No correlation was found with the number of phases involving stakeholders, but the number of stakeholder groups was correlated significantly (although not very strongly) with documentation (table 3). The aspects content, purpose, methodology, indicator scoring, indicator aggregation and interpretation of the results for which we asked about background documents roughly correspond with the 6 phases in ISA development. The associations between the individual types of documentation and the stakeholder types involved in the corresponding phase were rather disappointing though. Particularly farmers' participation and documentation availability show negative associations in all phases.

3.2.5. Implementation

ISAs seem to have a better chance of being implemented if they have multiple purposes and if more background documents are available: both show a rather weak, but significant correlation with implementation yes/no (table 3). The total number of indicators and the time needed for data collection in contrast do not seem to influence implementation, as no correlation was found.

Detailed association analysis shows:

- Implementation on project basis is associated with "other" purposes than the ones listed in the survey (consultancy, teaching, impact assessment and policy support were mentioned); various applying users (extension worker, researcher, civil servant, others except auditors); researchers or policy makers as end-users; and a wide availability of background documents.
- Commercial implementation is associated with the reporting purpose (+ 0.68), not with research (- 0.66); assessment at farm or industry-wide level, not landscape level (- 0,98); "other" end-users, such as "businesses, investors and banks" or "supply chain operators: food companies, retail, … up to consumers".
- Certification obviously is associated with the certification purpose, but not with research (the opposite of implementation on project basis); the farm as assessment level; auditors and sometimes farmers as applying users; farmers as end-users, as well as others (the buyers).
- Use by farmers is associated with farm-level assessment; civil servants as applying users; "other" users, as for most of the commercial or certification ISAs also "used by farmers" was ticked as implementation type. Surprisingly, it is NOT associated with farm development as a primary purpose; nor with the farmer as end-user of the ISA.

Participation by few stakeholder groups showed significant association with implementation as such. However, farmer participation was rather positively associated with most types of implementation, particularly with certification and use by farmers.

5. Preliminary conclusion

The survey of integrated sustainability assessment methods reached a 75 % response rate and resulted in an abundance of data on the ISA methods' characteristics, revealing a large variation between the ISAs in the survey. Strictly reductionist representations were rare, but holistic ones ranged from less than ten to hundreds of indicators. Next to farm development, other (combinations of) purposes were found; a wide range of end-users; a spectrum of data collection, processing and scoring methods; and variate methods to combine indicators into an ISA. Stakeholder involvement in ISA development was found common practice, especially in the early phases, defining the sustainability framework and selecting the indicators.

Correlation analysis revealed many associations between the ISA characteristics. However, the amount of detail explored by the tetrachoric correlations also resulted in an explosion of association measures, which hinders detecting the interesting ones. Therefore these associations cannot suffice to discern between the myriad of ISA methods. Further research is needed, starting with cluster analysis of ISA methods and their characteristics. It also seems interesting to expand the quantitative research with qualitative research, e.g. in-depth interviews with ISA developers, to grasp the full extent of reasoning behind ISA methods and the difficulties in their implementation. Thus decisive conclusions may be reached on *how sustainability frameworks, metrics and tools and their implementation can be enhanced to futureproof agricultural decision making at multiple levels and multiple scales*. For now, this first pilot activity managed to shed some light on the complexity of ISA methods and the variability in their characteristics.

References

Binder, C.R., Feola, G. & Steinberger, J.K. (2010). *Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture*. Environmental Impact Assessment Review 30: 71-81.

Bonnet, D.G. & Price, R.M. (2005). *Inferential Methods for the Tetrachoric Correlation Coefficient.* Journal of Educational and Behavioral Statistics 30 (2): 213-225.

Brundtland (1987) *Our Common Future*. Oslo: the World Commission on Environment and Development. United Nations, retrieved from http://www.un-documents.net/our-common-future.pdf.

Chedzoy, O.B. (2006). *Phi-Coefficient.* In: Encyclopedia of Statistical Sciences, Wiley, retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess1960.pub2/abstract.

Coteur, I., Marchand, F., Van Passel, S., Schader, C., Debruyne, L., Wustenberghs, H. & Keszthelysi, S. (2016). *Benchmarking sustainability farm performance at different levels and for different purposes: elucidating the state of the art*. 12th European IFSA Symposium, Harper Adams University, UK, 12-15 July 2016.

Ekström J. (2011) *The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate*. University of California, Department of Statistics Papers, retrieved from http://escholarship.org/uc/item/7qp4604r.

FAO (2013) *SAFA Sustainability Assessment of Food and Agriculture systems. Guidelines version 3.0*. Rome, Food and Agriculture Organization of the United Nations.

Gregory, P.J. & Kougioumoutzi, E. (2016). *TempAg: An international Research Consortium for Sustainable Agriculture in Temperate Regions*. 12th European IFSA Symposium, Harper Adams University, UK, 12-15 July 2016.

Marchand, F., Debruyne, L., Triste, L., Gerrard, C., Padel, S. & Lauwers, L. (2014). *Key characteristics for tool choice in indicator based sustainability assessment at farm level*. Ecology and Society 19 (3): 46, http://dx.doi.org/10.5751/ES-06876-190346.

Pope, J., Bond, A., Morrison-Saunders, A. & Retief, F. (2013). *Advancing the theory and practice of impact assessment: Setting the research agenda*. Environmental Impact Assessment Review 41: 1-9.

Schindler, J., Graef, F. & König, H.J. (2015). *Methods to assess farming sustainability in developing countries. A review*. Agronomy for Sustainable Development 35: 1043-1057.

Triste, L., Marchand, F., Debruyne L., Meul, M. & Lauwers, L. (2014) *Reflection on the development process of a sustainability assessment tool: learning from a Flemish case*. Ecology and Society 19 (3): 47, http://dx.doi.org/10.5751/ES-06789-190347

Acknowledgements
The authors are grateful to the TempAg network for kindly supporting this first Pilot Activity.