Environmental Impact Assessment Review 84 (2020) 106415

Contents lists available at ScienceDirect

Environmental Impact Assessment Review

journal homepage: www.elsevier.com/locate/eiar

A delphi-style approach for developing an integrated food/non-food system sustainability assessment tool

Samantha M. Mullender[a,1,*], Mignon Sandor[b], Andrea Pisanelli[c], Jerzy Kozyra[d], Robert Borek[d], Bhim B. Ghaley[e], Adrian Gliga[b], Moritz von Oppenkowski[f,4], Tim Roesler[f], Eldina Salkanovic[e], Jo Smith[a,2], Laurence G. Smith[a,3]

[a] The Organic Research Centre, Elm Farm, Hamstead Marshall, nr, Newbury RG20 0HR, United Kingdom
[b] University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Calea Manastur 3-5, 400372 Cluj-Napoca, Romania
[c] National Research Council, Institute of Research on Terrestrial Ecosystems, Via Marconi 2, 05010 Porano, Italy
[d] Institute of Soil Science and Plant Cultivation in Puławy – State Research Institute, Czartoryskich 8, 24-100 Pulawy, Poland
[e] Department of Plant and Environmental Sciences, University of Copenhagen, Højbakkegård Allé 30, 2630 Taastrup/Copenhagen, Denmark
[f] Phillipps-University Marburg, Deutschhausstraβe 10, 35032 Marburg, Germany

ARTICLE INFO

Keywords:
Delphi
Agroforestry
Participatory approach
Sustainability indicator
Sustainability assessment
Sustainability framework

ABSTRACT

Sustainability assessment is a complex field and its uptake amongst agricultural producers limited. Furthermore, the scope of current sustainability assessment tools does not extend to systems in which food production is integrated with production of non-food biomass (*e.g.* agroforestry). Participatory approaches to tool development offer a means to overcome the subjectivity of researcher-led tool design and thus the potential to increase relevance and engagement. In this work we develop a Delphi-style methodology as a means to produce a sustainability assessment tool suitable to assess and feedback on an integrated food/non-food system. Using a widely accepted agricultural sustainability framework and an existing farm sustainability assessment tool as a base, stakeholders were engaged with across six countries and multiple stakeholder groups to identify key indicators to be added to the tool. The methodology developed is described in detail, framed in the setting of this tool development process but providing a novel framework applicable to any situation where indicators must be developed for a complex issue of interest across multiple perspectives and stakeholder groups. Feedback and learning from the experience is provided. It was found that, contrary to some opinion, the inclusion of a face-to-face discussion round as part of the Delphi procedure provides a valuable means for information exchange and a move towards consensus amongst stakeholders. By using a 'snowball' approach to the in person discussions, it appears too that the loss of the voices of more socially retiring individuals can be avoided. Final levels of agreement vary substantially across the different areas of sustainability, with indicators in some areas (*e.g.* environmental integrity) proving much less controversial than others (*e.g.* social wellbeing). Despite this, the methodology effectively reaches a level of consensus amongst diverse stakeholders sufficient to guide the selection of sustainability indicators with a good level of confidence.

1. Introduction

The concept of sustainability is not a concrete one. Whilst attempts have been made to develop a universal framework for its assessment – the FAO's 'Sustainability Assessment of Food and Agriculture systems' (SAFA) for example – the appropriate indicators, the benchmarks against which to judge them and even the components important for a sustainable system remain very much determined by the developers' personal values and beliefs. It is perhaps for this reason, alongside the researcher-perspective of many assessments and frameworks, that despite large numbers of options, the voluntary application of sustainability assessments by producers is rare.

* Corresponding author.
E-mail address: s.mullender@cantab.net (S.M. Mullender).
[1] Permanent address: Anthesis Group, 9 Newtec Place, Oxford, OX4 1RE, United Kingdom.
[2] Permanent address: MV Agroecological Research Centre, Moinhos de Vento, 7750-217 Espirito Santo, Portugal.
[3] Permanent address: School of Agriculture, Food and Environment, Royal Agricultural University, Cirencester, GL7 6JS, United Kingdom.
[4] Permanent address: PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft, Moskauer Straße 19, 40227, Düsseldorf.

https://doi.org/10.1016/j.eiar.2020.106415
Received 2 August 2019; Received in revised form 26 March 2020; Accepted 8 May 2020

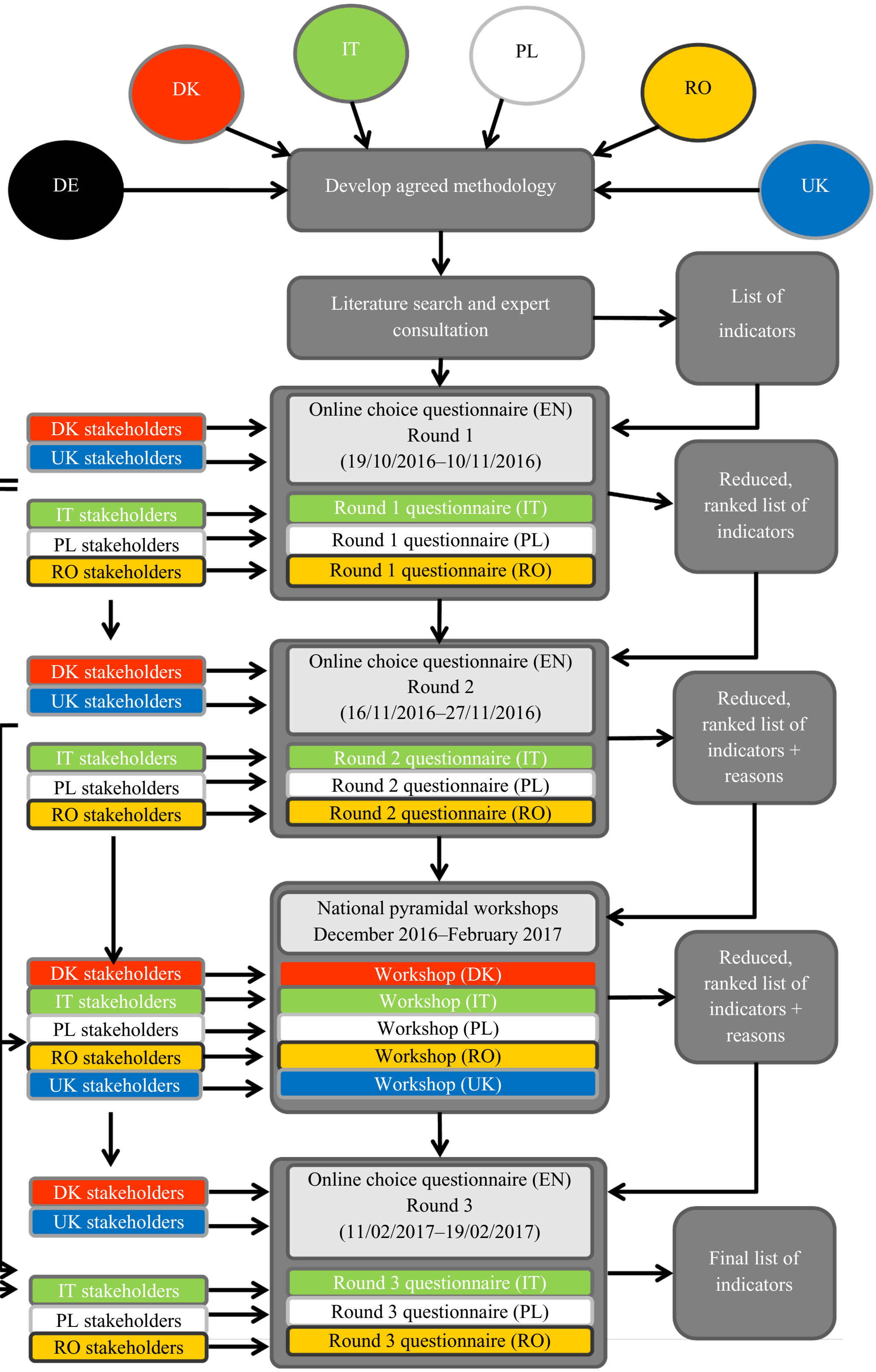

Fig. 1. Outline of research methodology. From an initial literature search to identify the indicator long list, three rounds of online questionnaire, offered in four languages, followed, separated by an in person workshop in five countries.
DK – University of Copenhagen; IT – CNR; PL – IUNG-PIB; DE – Phillipps-University Marburg; RO – UASVMCN; UK – Progressive Farming Trust.

One way to overcome the subjectivity caused by researcher decision-making in areas where rigorous scientific analysis or experimentation is not possible is to adopt a participatory or bottom-up approach to decision making (Delbecq et al., 1975 in Saint-Germain et al., 2000; Merfield et al., 2015; Schmitt et al., 2017, Rose et al., 2018). Participatory approaches to research are increasing, particularly in fields such as audits and assessments where the outcomes of the research will have to be adopted directly by stakeholders. Besides the belief that drawing on the knowledge of many increases the accuracy and relevance of the answer (Dalkey, 1969), there is evidence that adoption of a participatory approach increases subsequent perceived usefulness of the research outputs (deLancer Julnes, 2001).

A popular method for participatory research is the Delphi approach (Mukherjee et al., 2015). Whilst the exact implementation varies between projects, the approach can be identified by three integral principles – anonymous response, iteration and controlled feedback – and statistical group response. However, the Delphi approach as typically implemented poses a number of challenges, especially where the project and its stakeholders extend across multiple countries and represent stages of the value chain with potentially very different perspectives. As one example, the authority of the researcher can adversely affect the process, leading to bias in the wording of questions or the selection of experts (Avella, 2016). Researchers may similarly impose their preconceptions on the respondents, particularly where responses are sought to criteria identified in a literature review conducted by the same researcher or team (*ibid*, 2016). There can also be a tendency to become bogged down in discussions/debates over the method rather than the topic, whilst low response rates and inconsistent concept descriptions can lead to practical challenges with regard to data reliability and comparability (*ibid*, 2016).

This work describes the design and application of a Delphi-style approach for application in an international research project investigating the role of integrated food/non-food production in the design of more sustainable farming systems (SustainFARM – www.sustainfarm.eu). The method we develop, although applied here to sustainability indicators for agroforestry, provides a new way to integrate literature review, questionnaire-based Delphi and face-to-face Delphi approaches, in a way accessible to projects operating cross-border, cross-country and within a restricted time-frame and budget. The goal of the research presented is therefore to assess the suitability of this modified Delphi approach for achieving consensus on sustainability indicator selection within an international and multi-stakeholder research context.

Within the SustainFARM project, the Delphi approach was adopted to identify sustainability indicators for integrated production systems for use in the modification of an existing, agriculture-focused sustainability assessment tool. In this context, integrated food/non-food production was considered to include the use of waste residues from the agricultural system, as well as the use of incidental (*e.g.* hedgerows) or cultivated (*e.g.* tree rows in alley cropping) woody biomass, *i.e.* agroforestry.

The sustainability assessment tool selected for modification was the Public Goods tool (PG Tool, Gerrard et al., 2012), originally developed through two Defra-funded pilot projects that sought to combine data from scientific literature with expert-input in order to produce a framework for the evaluation of a farm's environmental, economic and social performance.

In the following section, we describe the modified-Delphi approach applied to select new indicators for the adaptation of the PG tool for agroforestry systems in Europe. We then go on to present the results of this exercise, both for the research itself and the performance of the method.

2. Methodology

Six national partners participated in this Delphi process: University of Copenhagen, Denmark – DK; CNR, Italy – IT; IUNG-PIB, Poland – PL; Phillipps-University Marburg – DE; UASVMCN, Romania – RO and Progressive Farming Trust (trading as The Organic Research Centre), United Kingdom – UK. Each contributed to different stages of the process shown in Fig. 1 below.

The first stage of the Delphi was to identify the pool of 'experts' whose knowledge would be drawn upon. Stakeholder categories were identified collaboratively with partner countries and comprised: farmer/land manager, local community, contractor, extension service/agricultural or forestry consultant, food and/or energy consumer, private company (energy supplier/renewable energy company, retailer of agricultural products, processor, *etc.*), environmental protection agency/conservation organisation, government body and researcher/academic. Whilst these were the best 'experts' in terms of assessing indicators for specialist Integrated Food/Non-Food Systems (IFNSs), it was also recognised that these were not all necessarily groups familiar with the fast-moving field of sustainability.

In the interests of promoting greater consensus, rather than brainstorming for new indicators, a literature search was therefore conducted to identify an initial list of indicators to feed a consultation. This took place over a period of c.four months (Oct. 2016–Feb. 2017) and comprised three online surveys with a practical workshop in each country between the second and third survey. The stakeholder list for this process was considered 'open', *i.e.* with the exception of the final round, stakeholders need not have been involved in previous rounds. This was to maximise the number and diversity of expert opinions represented. The final result was a list of the key indicators of sustainability for Integrated Food/Non-food Systems (IFNS) as identified by our stakeholders. The approach adopted is summarised in Fig. 1 and is described in detail below.

2.1. Literature search

A structured literature search was conducted by partners in the UK, Italy and Romania. IFNSs were divided into three categories: two types of 'traditional' system, where components are managed primarily for food but have the potential to provide non-food co-products, and one 'innovative' system, where food production is fully integrated with

Table 1
Suggested search terms to be applied in the selected search databases for each IFNS.

Crop/livestock system, making use of 'natural' incidental vegetation (e.g. valorising boundary hedgerows):		
"sustainab* indicator"	AND	hedge*
"economic indicator"		farming AND "wood fuel"
"environmental indicator"		agriculture AND "wood fuel"
"social indicator"		food AND "wood fuel"
"governance indicator"		"woody elements"
Crop/livestock system, making use of non-food residues/by products (e.g. using the pulp from olive processing):		
"sustainab* indicator"	AND	valoris* AND co-product AND farming
"economic indicator"		valoris* AND co-product AND agricultur*
"environmental indicator"		valoris* AND residue AND farming
"social indicator"		valoris* AND residue AND agricultur*
"governance indicator"		co-product AND farming
		co-product AND agricultur*
		residue AND farming
		residue AND agricultur*
Crop/livestock system integrated with woody components (agroforestry):		
"sustainab* indicator"	AND	biomass AND "food production"
"economic indicator"		biomass AND agricultur*
"environmental indicator"		biomass AND farming
"social indicator"		agro-forestry
"governance indicator"		integrated AND "short rotation coppice"
		"short rotation coppice" AND food
		"woody elements"

* indicates inclusion of all endings of given search term.

Table 2
Criteria and definitions given in the stakeholder survey for assessment of 'appropriateness'.

Criteria	Given definition
Relevance	Does the indicator have a logical or scientifically justifiable direct impact on the farm's sustainability?
Comprehensiveness	Is the indicator relevant to all systems combining food and non-food production (using hedges, using wastes, combining trees with crops/livestock)?
Interpretability	How easy is the indicator, and its consequences for farm management, to interpret?
Data quality	Can the indicator be measured accurately, either with existing data or through data collection?
Efficiency	Is data already available, or if not, is it quick and easy to collect?
Overlap	Is any similar indicator already in the PG tool (identified below the one presented) sufficient?

specific production of biomass for non-food uses. Suggested search terms were given for each category (Table 1) and translated into national languages. Searches were planned to continue until all search term/search database combinations were exhausted. However, fewer relevant results were found than anticipated, which, combined with time constraints, meant that only UASVMCN covered all the search terms.

Search engines and databases employed in the literature search included 1) Google (dominant search engine in all countries; returning grey and academic literature); 2) Google Scholar, or the best national equivalent (easy-to-access search for academic literature); 3) EuroStat; 4) other grey literature sources (EEA, FADN, OECD, UNECE, World Bank and national sources as appropriate). Only free to use search resources were used, thus excluding databases like Web of Science and Scopus. Due to their focus on international literature and lack of relevant resources in national language, search databases besides Google and Google Scholar were not explored in Italy and Romania.

Selected search terms were applied in every search database. The term, database and total number of results were recorded and the first 20 results (or all, if < 20 in total) manually filtered to remove results pre-dating the 1992 Rio Earth Summit (June 1992) or lacking 'reasonable reference' to an IFNS. The interpretation of 'reasonable' was left to the researcher but included sources that considered IFNSs without them being the focal topic. The first five of the filtered results (or all, if < 5 were relevant) were reviewed and all sustainability indicators contained within them – specific and non-specific to IFNSs – were recorded. Only indicators defined within the paper were included; no attempt to develop indicators based on the wider contents of the paper (beneficial qualities of the production system, for example) was made. Sources reviewed during previous searches were excluded from the five and not replaced. The same search term was used until 15 results had been reviewed or all databases searched, and a maximum limit of 50 reviewed results per IFNS category was set. The limits were set to ensure that wide coverage of the different forms of IFNSs and the different domains of sustainability was achieved within the search constraints.

The literature search was supplemented by consultation with the identified stakeholders from Denmark, Germany, Italy, Poland, Romania and the UK. Stakeholders were asked to identify websites, reports or other resources related to:

- Recommended management practices for IFNSs, in general or for a specific production system;
- Indicators of sustainability for these systems used by practitioners/assessors/others;
- The characteristics of a 'good' system (again, in general, or for a specific system).

The stakeholder-suggested resources were filtered and assessed in the same way as those from the literature search. The indicators identified from all three sources (literature search resources, stakeholder-suggested resources and stakeholder-suggested indicators) were compiled into a single list.

Finally, indicators were generalised and grouped to generate a list of unique, but not exclusive, indicators (*i.e.* the same indicator may exist at a general and more specific level – 'costs' and 'cost of feed', for example). These indicators were defined in accordance with the SAFA domains of sustainability: Good Governance, Environmental Integrity, Economic Resilience and Social Wellbeing (FAO, 2013). SAFA was selected as a framework in the light of its establishment to be a 'universal' framework for assessing agricultural sustainability.

As the sustainability assessment tool selected for use in the study (PG Tool) was designed for food production systems, it was taken that the relevant indicators for pure arable and/or livestock systems would already be present. The exception to this was for 'Good Governance' indicators, as this concept was, at the time, outside the scope of the PG Tool. All agriculture-relevant, non-governance indicators were therefore removed from the list, along with any other indicators already included in the tool. Indicators partially covered by more general, or a similar but non-identical indicator, already present in the tool were left in but presented alongside the related indicator(s). The indicators left after this refining were grouped according to subject and rephrased to be less abstract.

The resulting list was used in a pilot 'Delphi' process with five of the project partners (DK; IT; PL; RO and UK), comprising one online survey and a workshop where partners discussed in groups of eight to select their top indicators in each of the four SAFA sustainability domains. Feedback from this was used to further combine and/or remove similar indicators, adjust the order of the lists to present to stakeholders and elaborate certain indicators with definitions. The final method derived from this pilot is described below.

2.2. Online survey round 1

The survey was produced using the Qualtrics web platform (Qualtrics, 2017) and consisted of an introduction, a question identifying the stakeholder group(s) of the respondent and then the main body of the survey, comprising the indicator lists subdivided into the four SAFA domains of sustainability. For each list, participants were asked to select the five most 'appropriate' indicators for the assessment of farm-level sustainability of IFNSs, with particular regard to the indicators' relevance, comprehensiveness, interpretability, data quality, efficiency and overlap (see Table 2). They were reminded of these criteria and their definitions at the start of each of the four lists. Ranking was not used: the number of indicators in each list was too great for ranking to be meaningful. A tick box beside each indicator allowed respondents to indicate if they did not understand it. If ≥ 1/3 of the indicators within a domain were not understood, the respondent's responses within that domain were excluded. This was because:

a) it was felt that these respondents could not be considered 'experts';
b) there was a risk that they were rushing through the survey and just didn't want to think about their answers, making their other answers

unreliable; and

c) the indicators that they had answered for were at higher risk of being misinterpreted.

The survey was either translated into national languages by partners or circulated in English (Fig. 1) and was left open for 2–3 weeks to allow for any spread by 'word of mouth' to maximise the number of stakeholders that engaged.

2.3. Online survey round 2

The second survey used the preferences identified in the first round as feedback to encourage consensus. The number of votes received by each indicator was summed across countries and all indicators with no votes were removed. Besides the reduced lists, the second survey was of identical structure to the first with the following additions:

1. the number of votes received by each indicator was displayed alongside it; and
2. if a participant selected an indicator with ≤10 votes, there was a follow up question asking for their reasons/justifications for selecting an unpopular indicator.

Respondents were not warned beforehand that a reason would be required and were forced to give a reason by use of validation rules. Technical constraints, however, meant they were not reminded of their previous answers when completing the survey. The survey was available for one week.

2.4. Workshop

All survey respondents, along with some supplementary stakeholders, were invited to attend a practical workshop in their own partner country. Full, detailed instructions were provided to each country for the structure and organisation of this workshop. A target of 12 participants from across stakeholder categories was set, with a minimum number of three participants required to proceed.

Workshops began with an introduction, for which structured guidance was given. The instructed content comprised: 1. an introduction to the project, the concept of sustainability, the sustainability framework being used in this research (the FAO SAFA guidelines) and the three types of integrated food/non-food production being explored; 2. introduction to the sustainability assessment tool being adapted (the PG Tool); and 3. explanation of the Delphi process, the workshop's situation within it and how the workshop would proceed. Stakeholders were then given a handout containing the indicators from Round 2, less those that had received no votes, and the reasons for selection given by respondents of Round 2 who had selected an indicator with ≤10 votes in Round 1. The total number of votes received in Round 2 was presented alongside each indicator.

Workshops proceeded using a derivative of the snowball or pyramid method (Mccall et al., 2013) that was developed in the research project 'Rural Alliances' (Goldman, 2005; Rural Alliances, 2015). Stakeholders first identified their personal 'top five' indicators and recorded them on their handout. They were then paired with a stakeholder of similar background and asked to discuss, agree upon and record a mutual 'top five'. Recording answers discouraged denial or change of answers after discussion with others commenced. Pairs joined to form groups of four to agree their mutual 'top five' and then fours joined again to give eights. To conclude the process, everyone came together to present their group 'top five's and the reasons for their selections. Participants were finally given the opportunity to anonymously identify any one indicator to add and/or remove from the amalgamated groups' list. This was to ensure all views, including less confident personalities, were recognised. Reasons had to be given for additions/removals. Understanding of indicators was not checked at the workshop stage as any participant uncertainty would be clarified during discussion.

Two domains were considered at a time – Environmental Integrity with Social Wellbeing and Economic Resilience with Good Governance – and ten minutes given to reach a consensus at each stage. It was recognised at the outset that, if the number of attendants was insufficient, there might not be > 1 participant group left after the final pairing. In this case, the facilitating partner recorded the reasons of the single group. Where the number of participants was not a multiple of two/four/eight, a table was provided to give the 'pairing' patterns.

2.5. Online survey – Round 3

The final survey round allowed results from each country to be compiled and served to eliminate any error introduced by the face-to-face discussion. Indicators with no votes during the workshops were again removed and indicators presented alongside the number of votes they received and the reasons given for selection of low-scoring indicators. Indicators individually recommended to be added or removed were also shown, but clearly identified as individual recommendations. The survey followed the same structure as the previous two surveys but was sent only to stakeholders who had participated in one or more of the previous surveys or workshop. Like round two, it was available for a week's duration, although communication difficulties meant that the Polish survey was only available for two full days.

Upon closure of the final round survey, indicators with no votes were removed to give a final list of indicators. Those with only a single vote were highlighted as in need of further research before use in adaptation of the tool and those with ≤3 votes (15% of respondents) were separated as being of low priority. This list was the outcome of the stakeholder-led identification.

3. Results

127 stakeholders contributed their expertise to this research: 25 as part of the literature search/consultation phase and 102 during the Delphi process (Table 3). There was a high drop-out rate at all stages and a tendency for biased coverage of stakeholder categories within each country. However, overall, coverage of all stakeholder groups was achieved (Table 3).

3.1. Literature review

244 searches were conducted in total by the three countries, resulting in the review of 139 literature sources (Table 4). This was supplemented by 40 resources (of which 13 met the relevance criteria) and 42 indicators (30 unique indicators) recommended by stakeholders. From these, 649 unique indicators were identified, 103 of which were specific to non-food biomass or IFNSs. These had uneven coverage of the SAFA themes, with the environmental and economic domains being best represented and governance almost entirely neglected (Fig. 2). With the concept of 'governance' as a domain of sustainability being relatively new – the main source being the 2012 SAFA guidelines – it is perhaps unsurprising it is relatively unrecognised in past literature. To address the lack of recognition, the SAFA indicators themselves were used to populate 'governance'.

From this overall list, removal of agricultural and duplicate indicators and refining and grouping of others derived a list of 87 indicators for presentation to stakeholders in Round 1 of the survey.

Table 3
Number of participants at each stage in the Delphi process, subdivided by stakeholder category and country. W indicates the in-person workshop.

Country	Round	Farmer/ land manager	Local community	Contractor	Extension service/ consultant	Food/ energy consumer	Supplier / processor	Environmental organisation	Government body	Researcher / academic	Unknown	**Total**
DK	**1**	0	0	0	1	0	0	2	1	1	0	**5**
	2	1	0	0	0	0	0	1	0	1	0	**3**
	W	0	0	0	0	0	0	0	1	5	0	**6**
	3	0	0	0	0	0	0	0	0	1	0	**1**
IT	**1**	4	4	0	1	1	2	0	0	6	0	**10**
	2	3	2	0	2	2	1	0	0	4	0	**8**
	W	0	0	0	2	0	0	0	0	3	0	**5**
	3	0	1	0	0	1	0	0	0	4	0	**5**
PL	**1**	4	0	2	1	1	0	2	1	4	0	**10**
	2	1	0	0	0	0	0	1	0	2	0	**4**
	W	4	0	0	9	0	0	0	0	0	0	**13**
	3	0	0	0	0	0	0	1	0	2	0	**2**
RO	**1**	5	4	0	10	6	2	1	0	4	0	**15**
	2	2	3	0	4	3	1	0	0	1	0	**8**
	W	2	0	0	4	0	0	5	0	2	0	**13**
	3	0	2	0	5	1	1	1	0	2	0	**8**
UK	**1**	4	0	0	1	3	1	2	2	4	0	**11**
	2	1	0	0	1	2	1	1	1	3	0	**6**
	W	5	0	2	1	0	0	2	2	1	3	**18**
	3	4	2	1	1	2	0	2	1	1	0	**8**
Overall	**1**	**17**	**8**	**2**	**14**	**11**	**5**	**7**	**4**	**19**	**0**	**51**
	2	**8**	**5**	**0**	**7**	**7**	**3**	**3**	**1**	**11**	**0**	**29**
	W	**11**	**0**	**2**	**16**	**0**	**0**	**7**	**3**	**11**	**3**	**55**
	3	**4**	**5**	**1**	**6**	**4**	**1**	**4**	**1**	**10**	**0**	**24**

Table 4
Numerical breakdown of the literature sources reviewed. NS: national stakeholder-suggested sources.

	UK	IT	RO	PL	DE	DK	Total
No. searches conducted	**44**	**11**	**189**	**–**	**–**	**–**	**244**
- No. search terms	6	9	97	–	–	–	–
- No. search locations	9	2	2	–	–	–	–
No. sources reviewed – total	**67**	**33**	**35**	**2**	**0**	**0**	**139**
- Literature search	58	33	35	–	–	–	**126**
- NS sources	11	0	0	2	0	0	**13**

These were roughly evenly split across the four domains (Table A1 A).

3.2. Surveys and workshop

51 responses (6–15 per country) containing some element of prioritisation were received in Round 1, of which four were terminated prior to completion. These four were included for the SAFA domains that had been completed. No responses were removed for lack of understanding and comprehension was generally high, with only five respondents indicating understanding of < 85% indicators. Eighteen respondents went on to complete Round 2, supplemented by an additional 11 'new' stakeholders to bring to the total responses to 29. Representation of stakeholder groups was varied, with contractors and government bodies unrepresented and academics/researchers over-represented.

The workshops represented 55 stakeholders divided into nine groups, whilst the final Round 3 survey was completed by 24 stakeholders. Of these 24, one was incomplete and one indicated inadequate comprehension of the Environmental indicators (no understanding of 5 out of 13 indicators) and was thus excluded from this domain.

Results of each stage of the Delphi-style process with regards to votes received and indicator removal are shown in Appendix A (Table A1). The greatest reductions in indicators occurred from Round 1 to Round 2 and Round 2 to the Workshops. This corresponds to both addition of reasons for selection and lower participation (thus higher probability of '0' votes). By the final stage, little further adjustment of answers was seen, with just three indicators being removed.

The level of agreement in every round, however, remained highly variable between indicators and domain. Some indicators – 'Landscape diversity' for example – were selected by 22 out of 23 respondents in Round 3, whereas others received only a single vote. In total, a quarter of selected Social and Governance indicators in Round 3 received only 1–3 votes. The Environmental domain showed the highest level of stakeholder agreement, with half of selected indicators voted for by more than 50% of respondents. The lowest agreement was in Governance, where three-quarters of indicators were voted for by ≤37.5% of responses (≤9 people).

This concluding stage of the Delphi-style process gave a list of 54 potential indicators for use in the adaptation of the tool. Of these, four were highlighted as in need of further research (only one vote) and eight were identified as of lower priority (≤3 votes).

3.3. Feedback from partners on the methodology

Feedback on the applied Delphi-style process was obtained from all partners involved. This was collected *via* short questionnaires distributed by email. All partners reported difficulties surrounding stakeholders' understanding of the indicators, which was exacerbated by indicators appearing under multiple domains (DK, UK). In some cases this was due to terminology or unfamiliar perspectives, but in others it arose from more intrinsic cultural and systems differences between countries (PL). However, whilst apparent during the workshops, this uncertainty was less observable from the survey responses.

Reaching a consensus was reported as a challenge by participants in DK, IT and RO, due to reasons ranging from the replication of indicators in several domains (DK), the number and similarity of indicators (IT) and too large groups (RO). Participants in DK reported that the

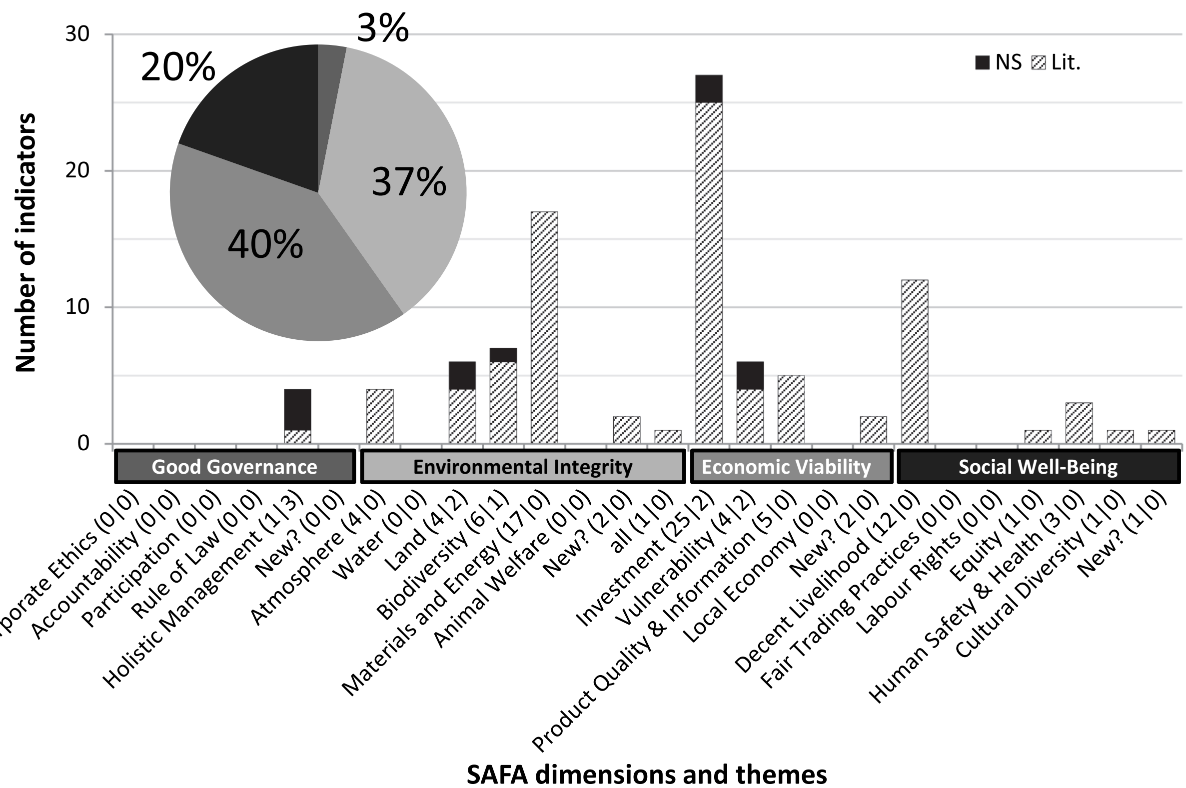

Fig. 2. Coverage of the SAFA four domains and 26 themes by the non-food/IFNS-specific indicators from the literature search sources (Lit.), although these may also have been covered by stakeholder-suggested sources, and from national stakeholder-suggested sources (NS). Totals are given in brackets adjacent to the theme, displayed as (Lit.|NS). Pie chart indicates proportional coverage of each domain (colours as per bar chart) for Literature and National Stakeholder indicators combined.

provision of reasons helped to encourage flexibility and movement towards consensus, as did having similar stakeholder backgrounds.

Another common feedback was that the time provided in the workshop outline was insufficient given the complexity of the subject. Participants in RO expressed that the time pressure meant they "moved away from the subject without any beneficial thing", whilst those in PL and RO felt more time was needed for introducing the general ideas of sustainability and IFNSs and emphasising the broader context besides economics and efficiency.

In general, the main suggestions for improvement were for fewer, more simplified indicators and more time spent on ensuring good understanding before asking for decisions. It was suggested that performing the exercise within the contextual framing of a specific IFNS might aid understanding.

4. Discussion

The conventional Delphi method has a number of limitations. Time and cost often makes it prohibitive from an organisational perspective (Bamberger and Mair, 1976), whilst for more complex or extensive data sets, retaining stakeholder engagement and engaging with more ambiguous issues can be a challenge. The method detailed in this study manages to address many of these issues, successfully reducing a long list of complex indicators to a much shorter list of indicators known to be relevant to stakeholders. The following sections reflect on the process applied, with particular regard to the three areas considered core to the difference between this process and what might be considered a 'conventional' Delphi: stakeholder engagement strategy, the inclusion of reasoning in the Delphi approach and the use of workshops. A final section discusses some of the other issues identified in the implementation.

4.1. Stakeholder engagement

Retention of participants is a recognised issue in Delphi procedures (Hsin-Ling et al., 2008, and references therein; Meijering and Tobi, 2016, and references therein), which can lead to problems with sample size and reliability of results. Allowing inclusion of new participants at all but the last stage – not typical in a Delphi approach – played an important role in maintaining a sufficient sample size throughout. By collecting reasons at each stage and sharing these with the following stage, it was possible to use the expertise of a dynamic pool of experts whilst ensuring that respondents were able to clearly judge the position of those who had been sampled already. Operating with open inclusion, this study was able to maintain sample sizes > 20 for all rounds. Although sample sizes smaller than this are often accepted in Delphi procedures (for example, Ogbeifun et al., 2016), inherent in a smaller sample is greater risk of inaccurate results. This is particularly true where the subject area is divisive and views of different nationalities and stakeholder groups differ, as was the case in this research. For an output useful to all groups upon completion, it was important to ensure that all perspectives were adequately represented. In a more linear methodology, the principle of saturation (Given, 2008) can be used to achieve this. This is where the researcher continues to survey until they stop getting new answers. However, where discussion and consensus is the focus, only surveying all potential participants beforehand or, as here, maximising the number and diversity of participants, can help attain a representative sample. An open pool of participants is therefore an effective solution to retaining the maximum possible sample size through a multi-stage, repetitive process.

4.2. Sharing reasoning for indicator selection

One of the problems that can occur when using a standard Delphi for complex, qualitative 'decisions' like agroforestry and sustainability, is that stakeholder groups have different interests and potentially limited understanding of other groups' perspectives. Furthermore, no one stakeholder will have a comprehensive knowledge of all research on the subject. This can then contribute to difficulty in reaching consensus or leave more scope for discussion to become dominated by stronger personalities. In this study we aimed to address this through provision of the reasoning behind answers from one round to the next.

Provision of reasoning is a debated principle in Delphi. The original RAND research found no evidence of any benefit and even potential detriment (Dalkey, 1969). However, recent thought tends to be that providing qualitative, in addition to quantitative, feedback is best practice. By drawing on the expertise of a group and making reasoning explicit, the method detailed above provided opportunity to educate and enlighten other participants. The multi-round process then allowed them to explore the alignment of the thoughts of others with their own thinking and give a more educated response in subsequent surveys. The effectiveness of providing reasons in this respect was evidenced in the Danish workshop, where it was explicitly fed back that provision of reasons helped with reaching consensus. With indicators subject to different interpretation (see '3.3 Feedback from partners on the methodology.'), reasons further benefited by revealing alternative interpretations to stakeholders and made it more difficult for researchers to exert any influence on the decisions of participants.

As a point of learning, in this study we presented reasons simply as a copy-paste of that given by the stakeholders. The disadvantage of this, however, was that reasons were in some cases sector or country specific. Without the surrounding context or background knowledge, the reason was meaningless. Although the workshops helped clarify ambiguous reasons, the benefit of this clarification would have been only intra-national and limited to workshop attendees. A simple, wider solution would have been for partners to translate contextually, rather than simply transcribe.

4.3. The structure of the Delphi process

There is a balance in any Delphi process between the number of rounds required to reach satisfactory level of consensus and participant fatigue at repeat questioning. There is also the question of format: digital or paper questionnaires *versus* in-person interaction and the presentation of the choices to be reviewed.

The four-round approach applied above strove to find a balance between stakeholder consensus on priority indicators and refinement to the point where important indicators were being lost. Few deletions from Round 1 to Round 2 reflects the diversity of the topic: with no information to promote consensus, limited agreement is a reasonable outcome. The limited change from the Workshop to Round 3, meanwhile, is then a likely indication that respondents are unwilling to adjust their position further and have reached a point of confidence in their own answers. This suggests that four rounds was a good number, and/or that the workshops were particularly helpful for consolidating shared opinion, although it is unknown what proportion of Round 3 participants had participated in a workshop. Although a further survey round would have confirmed whether stakeholders had indeed fixed their answers, feedback from the workshops suggests stakeholders were becoming impatient with the repeated questioning. A further round could therefore not only have risked unreliable answers, but also prompted disengagement with sustainability and future research. The four rounds, structured as two questionnaires, a workshop and a final questionnaire, seems an effective structure to balance consensus with participation and data quality.

A greater reduction in indicators from Round 2 to the Workshops and the Workshops to Round 3 would have been desirable. An easy way to achieve this would have been to adjust the boundaries – remove indicators with zero or one vote, for example, rather than just zero. The selected method for delimiting removal/provision of reasons was fairly arbitrary and it is possible to set the cut-off for removal/reasons using a number of more statistical approaches (receipt of a minimum proportion of the votes, retention of the 'n' most supported indicators *etc.*). Given the variation in participation between rounds, the former would be a more rigorous approach that would benefit in particular more quantitative studies. It would have also provided an opportunity to ensure all country and stakeholder perspectives were equally represented. With samples failing to achieve broad representation of all stakeholder categories and nationalities, there is a risk that under the approach above, the perspectives of some groups were lost through under-representation rather than low indicator quality. Weighting the votes of under-represented groups, or requiring all countries/stakeholder groups to show a lack of support in order for exclusion of an indicator, would avoid this.

The face-to-face workshop was included for two main reasons. First, to try to overcome the risks of misunderstanding that is prone to arise from lists of technical terms from diverse sectors, and second, to reduce the influence of the researcher on the final selection that could arise from the phrasing and grouping of the options presented in the questionnaire. A face-to-face element goes against many critiques of 'good Delphi practice', which state that in-person interaction can cause degradation of results (Graefe and Armstrong, 2011; Helmer-Hirschberg, 1967 and references therein) and lead to the views of less dominant personalities being lost (Boulkedid et al., 2011; Saint-Germain et al., 2000).

In multi-disciplinary scenarios like the one above, however, stakeholders are likely to represent a wide range of perspectives and areas of expertise and discussion may therefore be the only way that particular stakeholder groups can be made aware of the interests and knowledge of others. In the approach presented in this study, we show that there are methods that effectively reduce or even eliminate the risks associated with in-person discussion. This allows the benefits of more detailed conversation to be realised, leading to clarification of reasons and peer-to-peer learning (Hanafin et al., 2007).

In this study, two methods were applied to this end. Firstly, the snowball approach provides opportunity for participants to establish their view in small groups (one-to-one). As the group grows, a well-founded opinion will hopefully have already gained the support of others in the group. Secondly, participants were able to add indicators to the group selection or express discomfort with indicators selected anonymously. The fact that this option was provided but little utilised (just one addition and five removals were submitted out of 110 potential submissions) is support that individual views were not suppressed.

Besides the benefits of discussion, the workshops revealed universally a lack of understanding of the indicators and aims by some participants. This was not known from the online surveys, despite the inclusion of a 'don't understand' checkbox for every indicator. Besides promoting exchange of ideas, the workshops were thus extremely beneficial in increasing understanding for the final – and most influential – online survey. By adopting an adapted Delphi approach like the one here, it is therefore possible to use a face-to-face round to clarify differences in interpretation between stakeholder groups and overcome the difficulties in engaging diverse stakeholders in a multi-disciplinary consultation.

4.4. A side note on workshop structure and logistics

The main weakness of independent face-to-face workshops across countries is the slightly different research processes stakeholders then experience, despite the common set methodology. The advantage from the perspective of this study is that this has allowed the most vulnerable parts of the process to be identified for standardisation and the most effective workshop methodologies to be identified:

4.4.1. Workshop introduction

It was found that the workshop introduction was highly influential on what followed: the extent of stakeholder biases towards specific indicator areas (predominantly economics and efficiency); the framing of the proposed indicators in the context of specific integrated production systems; and purpose and function of indicators, for example. Language barriers made a single video recording to be played at all workshops impossible in this study, although this would be an option for geographically diverse intra-country workshops. However, a longer introduction and a more detailed script/recorded example for partners to all follow would have been beneficial.

4.4.2. Reaching agreement

All countries experienced difficulties in stakeholders reaching agreement within the set time limits. Reasons included misunderstanding of indicators and approach, difficulty reaching agreement in larger groups, too many/too similar indicators and crossovers between domains. Some extended the time until agreement was reached (RO), whilst others applied pressure to move on to the next phase (UK). Ensuring clearer understanding of indicators and approach beforehand could address most of the reasons given for difficulty and one way to assist this would be to require all workshop participants to have completed at least one prior round. However, given that in this case, many workshop participants were independent of previous rounds, there is a risk that this requirement would result in a substantially reduced number of participants and broadness of the stakeholder platform. More detailed exploration and participatory discussion in the workshop introduction offers an alternative solution. Timewise, a slight extension of the time limit is recommended; as Mullen (2003) highlights, consensus is desirable, but a false or a forced consensus is not. A limit, however, prompts decision, and thus ‘discussion until agreement’ is not recommended.

4.5. Other comments on the approach

Although the approach applied addresses many of the complexity and cost issues identified, it is possible that the simplified approach biased the selection of indicators towards those included within grey literature, in particular by the exclusion of the search engines Web of Science and Scopus in the initial literature review. Low response and high drop-out rates also affected the robustness of the study. The stakeholder groups consulted within the study could have driven the greater consensus within the domain of “Environmental Integrity”, as the indicators and the associated terms within this category may have been more familiar than those related to “Social Wellbeing” or “Good Governance”.

5. Conclusions

The Delphi process is a key research method where ambiguity and complexity persists, perspectives and interests differ and there is no ‘right’ answer. Yet what reflects ‘good practice’ in a Delphi is much debated and the method is currently associated with many limitations that make it hard to apply in any situation, but particularly in the circumstances of high complexity and diverse interests that it is best suited to.

This study takes one such area of research – sustainability indicators for integrated food/non-food production systems – and develops a Delphi-based approach that overcomes many of these limitations. The method developed successfully engages comparatively large numbers of stakeholders of multiple nationalities and interest groups over a multi-round engagement process, and it does this in a time- and cost-efficient manner. It recognises the need for clear understanding of others' perspectives and alternative interpretations of indicators – achievable only through sharing individual justifications/reasoning and through discussion. It also demonstrates the potential value of peer-to-peer learning that could arise as a repercussion of a Delphi approach, which it exploits through practical workshops and the sharing of reasons. However, it utilises these approaches in ways sensitive to their constraints and repercussions. The method applied still has room for improvement, some areas for which are highlighted in the discussion above. However, presented here is a method that is able to successfully prompt a move towards consensus amongst large numbers of diverse stakeholders on a controversial topic.

Acknowledgements

This work has been completed as part of the FACCE SURPLUS project ‘SustainFARM’ (http://www.sustainfarm.eu/). SustainFARM is a three year project, funded in the UK by Defra as part of the European FACCE SURPLUS ERA-NET co-fund programme formed in collaboration between the European Commission and a partnership of 15 countries. FACCE SURPLUS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652615. Thanks also go to the stakeholders, who gave up their time voluntarily to participate in the various stages of this research process.

Declaration of Competing Interest

None.

Appendix A. Indicator reduction process

Table A1
The presence/number of votes received for identified indicators at each stage of the Delphi-style approach. $(+/-\ n) = n$ individual suggested additions/removals; ▩ = indicator not presented at this stage. Number in brackets in heading indicates total number of participants at that stage.

Literature search output	Good Governance Indicators presented to stakeholders; presented order	R1 (/47)	R2 (/29)	W (/9)	R3 (/16)
	Application of recognised conservation methods	18	13	4	7
	Chain-of-custody control system *(i.e. creation/participation in a documented trail allowing a product to be traced along every stage of the supply chain, field to consumer)*	17	18	6	19
	Leakage effects *(i.e. displacement of any negativities associated with the old production system to a different geographic location, rather than stopping them from happening)*	7	2	0	
	Use of advisory services and associative organisations	16	8	2 (-1)	5
	Improvement of transparency of farm activities and policies *(i.e. availability, where appropriate, of farm strategies, goals, performance, decisions and decision-making processes to stakeholders)*	13	6	2	3
	Cooperation in supply chain (processors, retailers, etc.)	24	17	4	15
	Community (including consumers, local organisations, institutions and companies) participation	16	11	5	14
	Professional affiliations – strength and diversity (for example, cooperation with other producers)	7	1	0	
	Information/learning exchange	14	6	1	6
	Farm business reputation (moral and ethical)	8	0		
	Compatibility with national policy goals	4	0		
	Compliance with national laws and international agreements	5	4	1	2
	Farmer and employee training, including environmental	15	15	4	9
	Environmental checklist on employee training, waste management etc.	4	1	1	1
	Appropriate employer liability regulations *(i.e. the farm business' ability to compensate workers for illness or injury caused by their work on the farm)*	1	0		
	Property rights of farmer and previous land owner(s) *(i.e. whether the farmland was acquired in a legal and ethical way, and whether the current farmer's tenancy/ownership is secure)*	6	3	2	4
	Management complexity *(i.e. number of different levels and interactions within the farm management team: one farmer who owns and runs the farm would be simple; a conglomerate of farmers with an external farm manager would be more complex)*	4	3	1	2
	Sustainability management plan with clear objectives stated	8	4	3	9
	Monitoring of performance criteria	9	3	2	5
	Reproductive management plan *(i.e. existence of a managed breeding plan for livestock)*	1	0		
	Participation in agri-environment scheme(s)	2	1	0	
Sustainable forest management - crown density	Sustainable tree management - crown density **ALREADY IN PG TOOL**: Frequency that habitats are monitored and maintained (infrequently/regularly/etc.); extent that farm woodland is 'managed' (not at all/some/very active management for conservation/etc.)	2	2	0	
Sustainable forest management - planting density	Sustainable tree management - planting density **ALREADY IN PG TOOL**: Frequency that habitats are monitored and maintained (infrequently/regularly/etc.); extent that farm woodland is 'managed' (not at all/some/very active management for conservation/etc.)	0			
Sustainable forest management - tree height	Sustainable tree management - tree height **ALREADY IN PG TOOL:** Frequency that habitats are monitored and maintained (infrequently/regularly/etc.); extent that farm woodland is 'managed' (not at all/some/very active management for conservation/etc.)	0	1	0	
	Research and development programme in place	5	0		
	Use of 'Best Available Technology/Techniques' on farm	10	7	2	4
	Sustainability along the whole production chain *(i.e. whether processors/retailers/suppliers key to the farm activities are sustainable businesses)*	17	18	6	15

	Environmental Integrity	R1	R2	W	R3
Literature search output	**Actual prioritisation equivalent; presented order**	**(/51)**	**(/29)**	**(/9)**	**(/12)**
Accumulated leaf litter layer	Accumulated leaf litter layer *(i.e. amount of dead leaves on the ground)*	7	1	1	1
Tree litter					
Age distribution of trees	Age distribution of trees *(i.e. the mix of different ages of trees present)*	9	5	4	9
DBH distribution of trees					
Hedge length	Length of hedges (average and total)	10	4	0	
Shade protection - from weather, animal/insect pests, intruders, etc.	Shade protection - from weather, animal/insect pests, intruders, etc. *(i.e. amount of shade available to crops/livestock)*	10	5	1	3
Rooting depth	Rooting depth *(i.e. depth of the roots; of food and non-food plants)*	14	3	0	
Wind velocity - change due to non-food production	Wind speed - change due to non-food production	3	0		
Annual harvest of wood resources	Amount of non-food product harvested each year	17	7	0	
Game animals for sale of other products (horn, shell, bone etc.)	Sale of non-food products from game animals (horn, shell, bone etc.)	2	1	1	0
Grazed woodland historically grazed	Proportion of woodland grazed in the combined system that was grazed historically **ALREADY IN PG TOOL**: Woodland management for conservation/biodiversity (yes/no); Are livestock are excluded from woodland (yes/no)	10	5	1	7
Landscape diversity - agrarian ecotopica composition (crop ecotope composition)	Landscape diversity - agrarian composition (i.e. relative ratio of arboreal – tree based –harvested species and herbaceous harvested species)	32	25	8	22
Use of protected forests	Use of protected forests *(i.e. proportion of forest from which a harvest is taken that has a protected status)*	2	3	0	
Biodiversity - associated arboreal species	Change in biodiversity caused by adding a non-food product to a food system **ALREADY IN PG TOOL**: Rare species - number	20	13	4	13
Biodiversity - change due to non-food product					
Waste management - disposal of ashes	Disposal method for ash produced when burning the non-food product	0			
Energy balance - bioenergy feedstock processing	Energy balance – non-food processing and use *(i.e. amount of energy consumers can obtain from the non-food product compared to the amount of energy it contains in its raw form: a measure of whether it is better to harvest it, or leave it as part of the natural cycle)*	18	7	3	6
Energy balance - bioenergy use	Energy balance - non-food product *(i.e. the energy put into growing/harvesting/processing/etc. the non-food 'crop' versus the amount of energy made available to consumers: a measure of whether the non-food product in question*	23	16	6	13
Energy balance - non-food product					
Energy production (renewable) from agriculture and forestry					
	– biofuel for example – is better than currently used alternatives – in this case fossil fuels, wind power etc.)				
Fossil fuel use - change due to non-food product - cooling	Fossil fuel (coal, oil, gas, diesel, etc.) use - change due to non-food product *(i.e. replacement of fossil fuels on farm by non-food product and balance of any extra fossil fuels used for the production/harvest/etc. of the non-food crop).*	20	18	4 (-1)	13
Fossil fuel use - change due to non-food product - electricity					
Fossil fuel use - change due to non-food product - gaseous fuels					
Fossil fuel use - change due to non-food product - heating					
Fossil fuel use - change due to non-food product - liquid fuels					
Fossil fuel use - change due to non-food product - solid fuels					
Fossil fuel use - change due to non-food product - total					
GHG balance - non-food product	Greenhouse gas (GHG) balance - non-food product *(i.e. GHGs removed from the atmosphere by the non-food product vs the amount of GHGs released in its growth/processing/transport/use/etc.)*	20	13	6 (+1)	14
GHG balance - non-food product processing					
GHG balance - non-food product transportation					
GHG balance of non-food product/GHG balance of conventional use/non-use/disposal					
Energy density of biomass	Energy density of biomass *(i.e. if you burnt the non-food product, how much heat/electricity could you produce)*	15	6	3 (-1)	1
Feed conversion rate of non-food product	Feed conversion rate of non-food product *(i.e. if 1kg of the non-food product was fed to an animal, the amount of food product - milk/meat/eggs etc. - you would get out)*	10	3	0	
Processing efficiency - non-food feedstocks	Processing efficiency of non-food feedstocks *(i.e. a measure of the amount of the useful product lost when the non-food product is converted from its form on the field to the form in which it is sold)*	14	5	0	
COPIED ACROSS FROM SOCIAL IN R2	Land equivalence ratio/area equivalence index *(i.e. area of land that would be required to produce the same amount of food and non-food products as produced by the combined farming system in two separate farming systems)*	18	5	3	13
Land quality - change	REMOVED; NOT AN INDICATOR BUT A COMBINATION OF INDICATORS				

	Economic Resilience	R1	R2	W	R3
Literature search output	**Actual prioritisation equivalent; presented order**	**(/50)**	**(/29)**	**(/9)**	**(/14)**
Capital costs - non-food product	Capital costs - non-food product *(i.e. the initial cost of establishing production of the non-food product, e.g. purchase of specific harvesting machinery and planting new woody species)*	13	12	2 (-1)	5
Costs - processing non-food product		20	19	3	14
Costs - storage of non-food product	Total cost of non-food product from field to consumer *(i.e. costs associated with growing, harvesting, transporting, processing, storing, etc.)*				
Costs - transportation of non-food product					
Farm gate cost of non-food product feedstock/(costs of conventional use or non-use or disposal)	$\frac{\text{Farm gate cost of non} - \text{food product}}{\text{Costs of previous use/non} - \text{use/disposal}}$ *(i.e. changes to the costs to the farmer caused by the new use of the non-food product compared to the farmer's previous use of the biomass – for example, if the biomass was previously used as a fertiliser, the farmer may now need to buy in fertiliser)*	11	4	1	2
Cost of non-food product/cost of dedicated production	$\frac{\text{Cost of producing the non} - \text{food product in an integrated system}}{\text{Cost of dedicated production}}$ *(i.e. whether producing the non-food product in an integrated system is cheaper or more costly that a dedicated production system)*	11	6	3	10
Land equivalence ratio/area equivalence index	Land equivalence ratio/area equivalence index *(i.e. area of land that would be required to produce the same amount of food and non-food products as produced by the combined farming system in two separate farming systems)*	18	7	2	5
Demand for non-food product	Demand for non-food product *(i.e. how big is the potential market. For bioenergy, take into account the unused capacity available for its use)*	17	14	5	14
Capacity for bioenergy use/actual use					
Gross value added - non-food product	Gross value added - non-food product *(i.e. monetary value for the goods and services produced in relation to the non-food product, less the cost of all inputs, raw materials and negative environmental/social effects directly attributable to the non-food product)*	13	4	4	12
Revenue - non-food product	Revenue - non-food product *(i.e. total amount of money generated by the sale of the non-food product)*	11	4	1	0
Income - non-food product	Income - non-food product *(i.e. money made from non-food product after accounting for all associated financial costs)*	18	17	4	11
Income - sale of pruning materials					
Income - share of thinning					
Commercial extractive products (honey and nuts, medicinal herbs, regular fruits and vegetable supplements — wild leafy vegetables, roots, fruits)	Commercial extractive products (honey and nuts, medicinal herbs, regular fruits and vegetable supplements — wild leafy vegetables, roots, fruits) - total value harvested	11	4	3	9
Net present value - non-food product	Net present value - non-food product *(i.e. money made from non-food product minus costs associated with the non-food product over a period of time, but not accounting for time-related changes in monetary values - interest, for example)*	6	0		

Farm gate income of non-food feedstock/(income of conventional use or non-use or disposal)	$\frac{\text{Farm gate income from non}-\text{food product}}{\text{Income from previous use/non}-\text{use/disposal}}$ *(i.e. changes to farmer income caused by the new use of the non-food product)*	8	3	3	4
Added workload - non-food product Ease of residue recovery	Added workload - non-food product *(i.e. extra work required to turn the non-food product into a saleable product compared to if it was not used/cultivated)*	17	17	2	9
Harmonious production	Complementarity of production of food and non-food products (e.g. use same machinery for harvest; one provides shade/nutrition/other benefit to the other etc.)	17	16	9	19
Age distribution of trees Hedge - maturity	Age distribution of trees *(i.e. whether the trees are all the same age or a mixture of ages)*	1	0		
Sustainable rate of residue extraction	Sustainable harvest rate of residues/non-food crop *(i.e. proportion of 'wastes' left after harvest of the food product that can be removed whilst leaving sufficient amounts for soil fertility, land cover, social demands etc.)*	6	0		
Productivity - non-food feedstock Productivity - non-food product	Productivity - non-food product *(i.e. amount of non-food product in its usable/saleable form produced for a fixed level of inputs - area, labour, fertiliser, etc.)*	13	9	1	1
Harvest rate - trees	Harvest rate of trees/shrubs	1	0		
Yield - residues/sustainable recoverable yield of residue Theoretical yield - non-food product	Yield: $\frac{\text{amount of residues/non}-\text{food crop harvested}}{\text{sustainable harvest rate of residue/non}-\text{food crop}}$	4	1	1	0
Consistency of product composition - non-food product	Consistency of product composition - non-food product (e.g. energy content of wood fuel)	3	0		
DBH distribution of trees	Distribution of tree diameters *(i.e. do all trees have a similar sized trunk)*	0			
Oxidant formation potential of biofuel	Oxidant formation potential of biofuel from non-food product *(i.e. the tendency of a biofuel created as a non-food product to create oxidants - a type of air pollutant - when used)*	1	0		
Product contamination (non-food) - pesticide residues	Contamination of non-food product with pesticide residues	4	2	0	
Energy conversion efficiency of biomass	Energy conversion efficiency of biomass *(i.e. the proportion of the energy in the non-food substance that can be converted to a form - electricity, heat, etc. - that energy consumers can use)*	9	1	3	4
Feed conversion rate of non-food product	Feed conversion rate of non-food product *(i.e. if 1kg of the non-food product was fed to an animal, the amount of food product - milk/meat/eggs etc. - you would get out)*	14	5	0	
Growth rate - trees/shrubs	Time since last harvest – non-food resource (range and distribution)	4	0		
Hedge - top trimming	REMOVED; TOO INDIRECT				

Recoverability ratio of residue	REMOVED; SAME AS 'SUSTAINABLE RATE OF RESIDUE EXTRACTION'				
Reduced spending on otherwise purchased items	COMBINED WITH 'HARMONIOUS PRODUCTION'				
Time investment to harvest non-food product	REMOVED; SAME AS 'ADDED WORKLOAD - NON-FOOD PRODUCT'				
	Social Wellbeing	**R1**	**R2**	**W**	**R3**
Literature search output	**Actual prioritisation equivalent; presented order**	**(/49)**	**(/29)**	**(/9)**	**(/12)**
Added workload - non-food product	Added workload from new cultivation/use of non-food product *(i.e. the amount of extra time and effort required to produce the non-food product, compared to that required for its previous use)*	27	22	6	20
Average wage/projected average if non-integrated system	$\frac{\text{Average wage of farmer and workers}}{\text{Projected average wages under non} - \text{combined production systems}}$	30	21	6	20
Effect on access to energy - number of businesses		10			
Effect on access to energy - number of households	Effect on access to modern energy types (e.g. electricity, rather than wood fuel)		4	1	4
Effect on access to modern energy					
Effect on security of energy supply	Effect on security of energy supply *(i.e. the likelihood that demand for energy will exceed the capacity to supply it)*	22	17	7	17
Household time savings from modern bioenergy	Household time savings from non-food product (e.g. supply of modern bioenergy removing need to collect firewood)	16	8	1	6
Change in unpaid time spent by women and children collecting biomass	REMOVED; COVERED BY HOUSEHOLD TIME SAVINGS FROM MODERN BIOENERGY				
Alternative uses of harvested resources	Alternative uses of non-food product *(i.e. how the non-food elements now used to make the non-food product were used previously)*	15	7	2	3
Change in traditional biomass use due to non-food product	Changes to culturally traditional uses of the non-food feedstock due to new non-food product	13	2	1	3
Change in traditional biomass use due to non-food product	REMOVED; DUPLICATED UNDER THEME-LEVEL ANALYSIS				
Game animals for sale of other products (horn, shell, bone etc.)	Sale of non-food products from game animals (horn, shell, bone etc.)	6	0		
Food security - change in offer prices of national food basket					
Food security - effect of non-food production on food production	Food security - effect of combining food and non-food production on the amount of food produced at a field level *(i.e. the effects of interactions between the food and non-food crops on the yield of each other)*	31	25	7	20
Food security - effect of non-food production on food yield					
Food security - preference of marginal sites for energy crops	Preference shown for marginal sites for growing non-food crops *(i.e. extent that land quality is considered during its allocation to different crops/uses)*	17	6	3 (-1)	4
Product contamination (non-food) - pesticide residues	Contamination of the non-food product by pesticide residues	6	1	1	2
Permaculture	Application of permaculture principles (1. Provision for people to access the resources necessary to their existence; 2. Provision for all life systems to continue and multiply; and 3. Living within limits and distributing surplus)	27	15	5	12
Air pollution emissions of biofuel use/current average fuel emissions	$\frac{\text{Air pollution emissions of non} - \text{food product use}}{\text{Average emissions of products currently used for the same purpose}}$ *(for example, emissions from using the non-food product as a biofuel would be compared against emissions produced by the current fuel mix)*	25	17	4	8

References

Avella, J.R., 2016. Delphi panels: research design, procedures, advantages, and challenges. International Journal of Doctoral Studies 11 (1), 305–321 Doi:10.28945/3561.

Bamberger, I., Mair, L., 1976. Die Delphi-Methode in der Praxis: Ergebnisse einer explorationischen Untersuchung zu Einsatzbereichen und Anwendererfahrungen. Manag. Int. Rev. 16, 81–91.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., Alberti, C., 2011. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 6, e20476. https://doi.org/10.1371/journal.pone.0020476.

Dalkey, N.C., 1969. The Delphi Method: An Experimental Study of Group Opinion. Report no. RM-5888-PR. The Rand Corporation.

deLancer Julnes, P., 2001. Does participation increase perceptions of usefulness? An evaluation of a participatory approach to the development of performance measures. Public Perform. Manag. Rev. 24, 403. https://doi.org/10.2307/3381227.

Delbecq, A., Van de Ven, A., Gustafson, D., 1975. Group Techniques for Program Planning: a Guide to Nominal Group and Delphi Processes. Scott, Foresman & Co, Glenview, IL, USA.

Food and Agriculture Organization of the United Nations, 2013. Sustainability Assessment of Food and Agriculture Systems (SAFA) Guidelines. Food and Agriculture Organisation of the United Nations, Rome, Italy.

Gerrard, C.L., Smith, L.G., Pearce, B., Padel, S., Hitchings, R., Measures, M., Cooper, N., 2012. Public goods and farming. In: Lichtfouse, Eric (Ed.), Farming for food and water security. Springer Netherlands, pp. 1–22. https://doi.org/10.1007/978-94-007-4500-1_1.

Data saturation [online; accessed march 2020]. In: Given, L.M. (Ed.), SAGE Encycl. Qual. Res. Methods, https://doi.org/10.4135/9781412963909.n99.

Graefe, A., Armstrong, J.S., 2011. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. Int. J. Forecast. 27, 183–195. https://doi.org/10.1016/J.IJFORECAST.2010.05.004.

Hanafin, S., Brooks, A.-M., Carroll, E., Fitzgerald, E., GaBhainn, S.N., Sixsmith, J., 2007. Achieving consensus in developing a national set of child well-being indicators. Soc. Indic. Res. 80, 79–104. https://doi.org/10.1007/s11205-006-9022-1.

Helmer-Hirschberg, O., 1967. Analysis of the Future: The Delphi Method. Report no. P-3558. The RAND Corporation.

Hsin-Ling, H., Altschuld, J.W., Lee, Y.-F., 2008. Methodological and conceptual issues confronting a cross-country Delphi study of educational program evaluation. Eval. Program Plann. 31, 191–198. https://doi.org/10.1016/J.EVALPROGPLAN.2008.02.005.

Mccall, M., Ashley, H., Rambaldi, G., 2013. Tips for trainers: the snowball technique. Particip. Learn. Action 54 135–0. [online; accessed march 2020].

Meijering, J.V., Tobi, H., 2016. The effect of controlled opinion feedback on Delphi features: mixed messages from a real-world Delphi experiment. Technol. Forecast. Soc. Change 103, 166–173. https://doi.org/10.1016/J.TECHFORE.2015.11.008.

Merfield, C., Moller, H., Manhire, J., Rosin, C., Norton, S., Carey, P., Hunt, L., Reid, J., Fairweather, J., Benge, J., Le Quellec, I., Campbell, H., Lucock, D., Saunders, C., Macleod, C., Barber, A., Mccarthy, A., 2015. Are organic standards sufficient to ensure sustainable agriculture? Lessons from New Zealand's ARGOS and sustainability dashboard projects. Sustain. Agric. Res. 4 (3). https://doi.org/10.5539/sar.v4n3p158.

Mukherjee, N., Hugé, J., Sutherland, W.J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., Koedam, N., 2015. The Delphi technique in ecology and biological conservation: applications and guidelines. Methods Ecol. Evol. 6, 1097–1109. https://doi.org/10.1111/2041-210X.12387.

Mullen, P.M., 2003. Delphi: myths and reality. J. Health Organ. Manag. 17, 37–52. https://doi.org/10.1108/14777260310469319.
Ogbeifun, E., Agwa-Ejon, J., Mbohwa, C. and Pretorius, J. H. C., 2016. The Delphi technique: A credible research methodology. Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, March 8–10, 2016.
Rose, D.C., Parker, C., Fodey, J., Park, C., Sutherland, W.J., Dicks, L.V., 2018. Involving stakeholders in agricultural decision support systems: improving user-centred design. Int. J. Agric. Manag. 6 (3–4), 80–89. https://doi.org/10.5836/ijam/2017-06-80.
Rural Alliances, 2015. URL. http://www.rural-alliances.eu/ (online; accessed February 2017).
Saint-Germain, M.A., Ostrowski, J.W., Dede, M.J., 2000. Oracles in the ether: using an E-mail Delphi to revise an MPA curriculum. J. Public Aff. Educ. 6, 161–172. https://doi.org/10.2307/40215485.
Schmitt, E., Galli, F., Menozzi, D., Maye, D., Touzard, J.-M., Marescotti, A., Six, J., Brunori, G., 2017. Comparing the sustainability of local and global food products in Europe. J. Clean. Prod. 165, 346–359. https://doi.org/10.1016/J.JCLEPRO.2017.07.039.
Qualtrics, 2017. Qualtrics software. Copyright © 2017 Qualtrics. https://www.qualtrics.com.