



D3.1

Review of existing monitoring approaches for schemes and labels

ECOS

30 November 2022



PROJECT INFORMATION

PROGRAMME	Horizon Europe
TOPIC	HORIZON-CL6-2021-ZEROPOLLUTION-01-07
TYPE OF ACTION	HORIZON-CSA
PROJECT NUMBER	101059785
START DAY	1 June 2022
DURATION	36 months

DOCUMENT INFORMATION

TITLE	Review of existing monitoring approaches for schemes and labels
WORK PACKAGE	WP3 Monitoring system and testing on schemes and labels
TASK	Task 3.1 Review of available monitoring approaches and indicators
AUTHORS (Organisation)	Mariana López Dávila (ECOS)
REVIEWERS	Iris Vural Gursel (WR), Katrin Wingert (CU), Loek Verwijst (CU), and Mathilde Crêpy (ECOS)
DATE	30 November 2022

DISSEMINATION LEVEL

PU	Public, fully open	x
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444	
Classified S-UE/EU-S	EU SECRET under the Commission Decision No2015/444	

DOCUMENT HISTORY

Version	Date	Changes	Responsible partner
0.1	11.11.2022	First draft ready for review	ECOS
0.2	16.11.2022	Comments/edits from reviewers provided	WR, CU, ECOS
0.3	22.11.2022	Final draft submitted to the coordinator	ECOS
1.0	25.11.2022	Deliverable finalized for submission	WR
1.0	30.11.2022	Submission	WR

LEGAL NOTICE

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

© **SUSTCERT4BIOBASED Consortium, 2022**

Reproduction is authorised provided the source is acknowledged.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	8
1. INTRODUCTION.....	9
2. METHODOLOGY	9
2.1 Phase 1: Scanning	9
2.2 Phase 2: Screening.....	10
2.3 Phase 3: Synthesizing	12
2.3.1 <i>Review and synthesis of criteria</i>	12
2.3.2 <i>Review and synthesis of rating and scoring methodologies, visualisation of results.....</i>	14
3. FINDINGS	14
3.1 Overview of the selected tools	14
3.1.1 <i>Guidelines</i>	14
3.1.2 <i>Assessment tools</i>	15
3.1.3 <i>Benchmarking tools.....</i>	15
3.1.4 <i>Evaluations.....</i>	16
3.1.5 <i>Legislation</i>	16
3.2 Synthesis of evaluation criteria of shortlisted tools	17
3.2.1 <i>Theme 1: Scheme governance.....</i>	18
3.2.2 <i>Theme 2: Scheme management</i>	19
3.3 Summary of rating and scoring methodologies	21
3.3.1 <i>Certification Assessment Tool (CAT).....</i>	21
3.3.2 <i>Farm Sustainability Assessment (FSA)</i>	22
3.3.3 <i>Sustainability Assessment in Food and Agriculture (SAFA).....</i>	24
3.3.4 <i>Sustainability Standards Comparison Tool (SSCT)</i>	26
3.3.5 <i>Benchmarking Platform (BP)</i>	27
3.3.6 <i>Evaluation of Swiss Food Labels (ESFL).....</i>	27
3.3.7 <i>Sustainability Monitoring and Assessment RouTine (SMART).....</i>	29
4. DISCUSSION AND CONCLUSIONS	30
5. REFERENCES.....	32
ANNEX A: CRITERIA SYNTHESIS MATRIX	34

TABLE OF FIGURES

Figure 1. Results of Tool Screening Process (Shortlists).....	12
Figure 2. Overview of Themes and Subthemes	18
Figure 3. CAT Scoring System [26].....	21
Figure 4. CAT Results Visualisation [26].....	22
Figure 5. FSA Self-Assessment Questionnaire [7].....	23
Figure 6. FSA SQA Dashboard [7].....	24
Figure 7. The SAFA Sustainability Dimensions and Themes [9]	24
Figure 8. SAFA Rating Scale [9].....	25
Figure 9. Visualisation of SAFA Results [9].....	25
Figure 10. Example of SSCT Scoring System [27]	26
Figure 11. SSCT Calculation Sheet Example [27]	27
Figure 12. SSCT Final Evaluation Result Example [27].....	27
Figure 13. BP Proposed Performance Level Scoring System [12]	27
Figure 14. ESFL Scoring Template [15]	28
Figure 15. ESFL Results Visualisation [15]	29
Figure 16. SMART Results Visualisation [16].....	29
Figure 17. SMART Summary Sheet [16].....	30
Figure 18, Initial classification of the how the criteria reviewed could be leveraged to measure effectiveness or robustness under our monitoring system	31

LIST OF TABLES

Table 1. Results of Tool Screening Process (Tools by Screening Criteria)	11
Table 2. Overview of the Purpose of the Tools Reviewed	17
Table 3. ESFL Maximum and Weighted Scores	28

ABBREVIATIONS

BP	Benchmarking Platform
CAT	Certification Assessment Tool
CGF	Consumer Goods Forum
EC	European Commission
EPA	Environmental Protection Agency
EPPP	Environmentally Preferable Purchasing Program
ESFL	Evaluation of Swiss Food Labels
FAO	Food and Agriculture Organization
FSA	Farm Sustainability Assessment
GBT	Global Benchmarking Tool
GIZ	German Corporation for International Cooperation
GSSI	Global Sustainable Seafood Initiative
HLS	Harvard Law School
ISO	International Standardisation Organisation
M&E	Monitoring and Evaluation
MSI	Multi-Stakeholder Initiatives
RED II	Revised Renewable Energy Directive
SAFA	Sustainability Assessment in Food and Agriculture
SAI	Sustainable Agriculture Initiative
SAQ	Self-Assessment Questionnaire
SFS	Sustainable Food Systems
SMART	Sustainability Monitoring and Assessment Routine
SSCI	Sustainable Supply Chain Initiative

SSCT	Sustainability Standards Comparison Tool
STAR-ProBio	Sustainability Transition Assessment and Research of Bio-based Products
UNCTAD	United Nations Conference on Trade and Development
VSS	Voluntary Sustainability Standards
WWF	World Wildlife Fund

Executive Summary

The goal of this deliverable was to conduct an in-depth review of existing guidelines, assessment tools, benchmarks, evaluations, and legislations (“tools”) that assess sustainability certification schemes and labels. This review was used to identify existing criteria, rating and scoring methodologies, and visualisation approaches that could be leveraged by the project team when developing its monitoring system. The authors reviewed the criteria of 13 tools, and the rating and scoring methodologies and visualisation of results of 7 tools. The review of the criteria of all the tools combined was synthesized into a list of 137 criteria, 18 subthemes, and 2 themes. The review of rating and scoring methodologies and visualisation approaches resulted in detailed summary descriptions for each of the tools.

The study concluded that many of the criteria identified in this review is adequate for measuring either effectiveness or robustness and therefore can be leveraged to design our monitoring system, and that when developing criteria, the team will create different tiers (i.e. basic/advanced). Moreover, preliminary thoughts on the level of complexity of the rating and scoring methodology as well as ideas on the visualisation of the results are elaborated on. In sum, this detailed overview provides valuable input in the development of our monitoring system in accordance with the goal of our SUSTCERT4BIOBASED project.

1. Introduction

The following report presents the results of an in-depth review of existing guidelines, assessment tools, benchmarks, evaluations, and legislations (“tools”) that assess sustainability certification schemes and labels.

The review was performed with the goal to identify existing criteria, rating and scoring methodologies, and visualisation approaches that could be used as input in the design of our own monitoring system. To achieve its goal, the author conducted a detailed review of relevant 15 tools.

This report is organised as follows: section 2 details the methodology used to select and analyse the tools that were reviewed, section 3 presents the findings of the review, and section 4 presents the discussion and conclusions of this exercise.

2. Methodology

The goal of this review was to identify existing criteria, rating and scoring methodologies, and visualisation approaches that could be leveraged by the project team when developing its own monitoring system. This goal was achieved through an in-depth document review. The document review was conducted using a multiphase process which involved 1) scanning, 2) screening and 3) synthesizing. The following section details each of the three phases.

2.1 Phase 1: Scanning

The scanning phase consisted of identifying all potentially relevant documents through a systematic online search for guidelines, tools, evaluations, benchmarks, legislation, and other sources that contained criteria and methodologies to evaluate sustainability certification schemes and labels. Relevant documents and projects were identified using Google Search. The search was conducted through the combination of keywords using Boolean operators (AND/OR):

- *{sustainability certification, certification scheme, sustainability standard, sustainability systems, voluntary sustainability standard, ecolabel}*
- *AND {tool, assessment, monitoring tool, evaluation, minimum criteria, minimum requirements, guidelines, best practices, benchmark, and benchmarking tool}*
- *OR {database, map, repository, compilation}*.

In addition, the author received suggestions from SUSTCERT4BIOBASED team members from Control Union and Wageningen Research. As a result, eighteen relevant documents were identified. These documents were a combination of guidelines (4), assessment tools (7), benchmarking tools (3), evaluations (2), and legislations (2). A detailed list is presented below. For simplicity, this document will use the word “tool” to refer to any of the documents listed below regardless of whether it is a guideline, assessment tool, legislation or evaluation.

Guidelines

1. Standard-Setting Code of Good Practice Version 6.0 by ISEAL Alliance [1]
2. Assurance Code of Good Practice Version 2.0 by ISEAL Alliance [2]
3. Impacts Code of Good Practice Version 2.0 by ISEAL Alliance [3]
4. Benchmarking Criteria Version 2.0 by ISEAL Alliance [4]

Assessment Tools

5. Butterfly Model by LETZ CERTI [5]
6. Certification Assessment Tool (CAT) Version 4.0 by World Wildlife Fund (WWF) [6]
7. Farm Sustainability Assessment (FSA) Version 3.0 by the Sustainable Agriculture Initiative (SAI) Platform [7]
8. Multi-Stakeholder Initiatives (MSI) Evaluation Tool Version 1.0 by International Human Rights Clinic at Harvard Law School (HLS) [8]
9. Sustainability Assessment in Food and Agriculture (SAFA) Version 3.0 by the Food and Agriculture Organization (FAO) [9]
10. Sustainability Standards Comparison Tool (SSCT) Version 2.0 by German Corporation for International Cooperation (GIZ) [10]
11. Voluntary Sustainability Standards (VSS) Assessment Toolkit by United Nations Conference on Trade and Development (UNCTAD) [11]

Benchmarking Tools

12. Benchmarking Platform (BP) for Certification Schemes by the Sustainability Transition Assessment and Research of Bio-based Products (STAR-ProBio) project [12]
13. Global Benchmarking Tool (GBT) Version 2.0 by the Global Sustainable Seafood Initiative (GSSI) [13]
14. Sustainable Supply Chain Initiative (SSCI) Version 1.1 by Consumer Goods Forum (CGF)[14]

Evaluations

15. Evaluation of Swiss Food Labels (ESFL) by the Pusch Foundation [15]
16. Sustainability Monitoring and Assessment RouTine (SMART) Sustainability Assessment by Sustainable Food Systems (SFS) [16]

Legislation

17. Environmentally Preferable Purchasing Program (EPPP) Assessment Framework Version 2.0 by the United States Environmental Protection Agency (EPA) [17]
18. Revised Renewable Energy Directive (RED II) Sustainability Criteria by European Commission (EC) [18]

2.2 Phase 2: Screening

The screening phase consisted of a high-level review of all the tools identified in the scanning phase to determine which ones would be included for further review. The author developed two screening criteria to guide this phase.

Criteria 1. First, since this report is meant to draw insights into how to assess the performance of certification schemes, the tools were screened to select only those dealing with certification schemes and labels. Sixteen of the 18 schemes were selected. The two schemes that were not selected dealt with the evaluation of farms (FSA and SMART) which was not relevant for this study. Next, the author selected only those tools which evaluated the effectiveness, robustness, or overall sustainability performance of certification schemes and labels. This step narrowed the number of tools from 16 to 13. The tools that were not selected were considered irrelevant because their criteria assessed other aspects such as the adoption of VSS schemes (VSS), human rights (MSI), and market value (Butterfly Model).

Criteria 2. Second, the author conducted a high-level review of the scoring and rating methodologies of the 18 tools. The tools that had a publicly accessible scoring and rating

methodologies were selected for further study. This screening step led to the identification of 7 tools. Lastly, the author reviewed the 18 tools looking for visualisation of results. Coincidentally the same 7 tools were selected for review of their results visualisations.

Table 1 below, shows the results of the screening process.

Table 1. Results of Tool Screening Process (Tools by Screening Criteria)

	Criteria 1		Criteria 2
	Does the tool assess certification schemes and labels?	Do the criteria evaluate the effectiveness, robustness or sustainability of certification schemes and labels?	Does the tool have a public scoring and rating methodology? Does it visually present results?
Guidelines			
1. Standard-Setting Code, ISEAL	Yes	Yes	No
2. Assurance Code, ISEAL	Yes	Yes	No
3. Impacts Code, ISEAL	Yes	Yes	No
4. Benchmarking Criteria, ISEAL	Yes	Yes	No
Assessment Tools			
5. Butterfly Model, LETZ CERTI	Yes	No	No
6. CAT, WWF	Yes	Yes	Yes
7. FSA, SAI Platform	No	No	Yes
8. MSI, HSL	Yes	No	No
9. SAFA, FAO	Yes	Yes	Yes
10. SSCT, GIZ	Yes	Yes	Yes
11. VSS, UNCTAD	Yes	No	No
Benchmarking Tools			
12. BP, STAR-ProBio	Yes	Yes	Yes
13. GBT, GSSI	Yes	Yes	No
14. SSCI, CGF	Yes	Yes	No
Evaluations			
15. ESFL, Pusch	Yes	Yes	Yes
16. SMART, SFS	No	No	Yes
Legislation			
17. EPPP, US EPA	Yes	Yes	No
18. RED II, EC	Yes	Yes	No

The screening process resulted in the selection of 15 tools which were organized into three shortlists: tools to be reviewed for their criteria (8), tools that would be reviewed for their criteria, rating, scoring and visualisation (5), and tools that would be reviewed for their rating, scoring and visualisation (2). Figure 1 below depicts these results visually (note that the numbering in Figure 1 corresponds with the numbering in Table 1 and in the list of tools provided in section 2.1).



Figure 1. Results of Tool Screening Process (Shortlists)

2.3 Phase 3: Synthesizing

The final phase consisted in conducting an in-depth review of the criteria, rating and scoring methodologies, and visualisation of results of the shortlisted tools.

2.3.1 Review and synthesis of criteria

The process to review and aggregate all the criteria covered in the shortlisted tools was performed in three steps.

Step 1. Criteria compilation. The authors began by compiling all the criteria in an Excel document, tagging each criterion to its designated sub-theme, theme, and tool. From this initial exercise, the author produced a criteria overview table, which contained the name of the tool in the first cell a column followed by all the criteria listed in the subsequent rows on that column. The overview table allowed the author to see the individual criteria for each tool side by side. The criteria from the four ISEAL tools were listed under a single column to reduce the number of columns and facilitate the review process.

During this exercise, sector specific criteria (i.e. agriculture, forestry, seafood, etc.) and criteria related to environmental, social and/or economic objectives were not added to the Excel document. The reason for this omission was to avoid overlap with the Task 1.2 which compiled environmental, social, and economic criteria reviewing standards, legislation, sustainability certification schemes/ ecolabels and studies.

Step 2. Criteria merging. Next, the author created a new column in the criteria overview table to list all the criteria mentioned in all the tools only once. The author then copy-pasted the ISEAL column into this new column to use ISEAL's criteria as the starting point for this exercise. The choice to use ISEAL's criteria as the starting point was deemed appropriate given that it was found that most of the tools relied on ISEAL's Codes of Good Practice as the basis for the development of their criteria.

Therefore, the wording of the criteria in this review (see Annex A) heavily resembles ISEAL's criteria. However, the language is not identical since the author made modifications when deemed necessary to better illustrate the content of the criteria in review.

The next step, given that the ISEAL column combined criteria of four different ISEAL tools, was to delete any repetitions to clean this new column and ensure each criterion was unique.

Once the column was clean, the author went through each criterion for each of the tools in the criteria overview table. Every time a criterion matched a criterion in the new column, this criterion was highlighted to symbolise that it was included in the new column. When the criterion did not match any of the criteria in the new column, that criterion was included in the new column. In this way, the author ensured that all the criteria were listed under a single column.

This exercise produced a total of 137 criteria which reflects criteria included in all the tools combined. Although there is significant overlap across the criteria in the tools reviewed, it is important to highlight that not all 137 criteria are included in all the tools. Annex A illustrates this by showing which of the 137 criteria is included in each of the tools reviewed.

It is also important to highlight that the goal of this merging exercise was to aggregate all the criteria with a sufficient level of detail to provide an in-depth overview of the criteria themselves. However, it is not an exercise meant to include every single detail of every single tool. This means that there was some level of nuance and detail lost in the merging exercise, for example, in some cases the same requirement was presented with more detail in one tool than another. In most of these cases, the criteria was kept at a general level, unless the added detailed significantly changed the requirement, in that case a new criteria was added.

In order cases, some tools had requirements written in broader ways but were accompanied by very detailed guidelines, while other tools use detailed requirements instead of detailed guidelines. The details included in the guidelines were not included in this review since this exercise did not drill beyond the criterion level to keep the scope of this review to a manageable level. The result of the exclusion of the application guidance is that, in Annex A, some criterion will not be marked as included under a specific tool when it could be the case that the tool does address that criterion through its application guidance (or somewhere else).

It is also worth emphasizing that the methodological choices employed to aggregate and merge the criteria in the tools that were reviewed are insufficient to perform a comparative analysis or an assessment of the tools themselves. Therefore, the findings produced in this report should not be interpreted as a benchmarking or evaluation exercise. The findings produced in this report should be solely used to understand what type of criteria are covered by the shortlisted tools.

Step 3. Criteria categorization. Once all the criteria were aggregated into one column, the author proceeded to organize the criteria into themes. All the reviewed tools organise their criteria around themes and subthemes. The number and level of themes and subthemes in the tools reviewed vary greatly, ranging from 3 to 20 themes, and from 0 to 62 subthemes.

The choice of the number of themes and subthemes is important since they play an important role in aiding the reader to process the criteria. Themes and subthemes are also an opportunity to emphasise a theme of particular importance to the author.

For this review, the goal of organising the criteria into themes and subthemes is for the reader to get a bird's eye view of all the types of criteria covered in the tools. Too few themes/subthemes will not provide that overview, while too many will become too difficult to process at one glance. The author, therefore, endeavoured to develop enough themes and subthemes to capture the breath of the criteria in one glance.

To select the categorisation scheme for this review, the author began by reviewing all the existing categorisations of the tools reviewed. During this first review, two overarching themes emerged: 1) criteria that assessed the overall governance of the scheme, and 2) criteria that assessed the management of the schemes. Having identified the two key themes, the author conducted a second review of the criteria under each of the two overarching themes. At this point, the author leveraged the existing classifications of the different tools to produce subthemes.

Under the first theme “scheme governance”, the author identified four subthemes:

1. General
2. Sustainability outcomes
3. Monitoring scheme performance
4. Transparency

Under the second theme “scheme management” the author identified 14 subthemes:

1. Standard availability and accessibility
2. Multi-stakeholder engagement
3. Decision-making
4. Assurance
5. Oversight
6. Accreditation
7. Certification
8. Group assessment
9. Competence
10. Impartiality
11. Chain of custody
12. Claims and labels
13. Consistent interpretation
14. Review and revision

2.3.2 *Review and synthesis of rating and scoring methodologies, visualisation of results*

This last part of the exercise consisted of a simple yet in-depth review of the rating and scoring methodologies, and of the visualisation of results of the shortlisted tools. The review resulted in short descriptions of the scoring procedures of each tool.

3. Findings

The following section presents the findings of this in-depth review. It begins with an overview of the shortlisted tools (section 3.1), then with the synthesis of the criteria (section 3.2) and finally it presents brief descriptions of the rating and scoring methodologies and results visualisation of the shortlisted tools (section 3.3).

3.1 Overview of the selected tools

As mentioned in section 2.2, this study selected 15 tools for review. These tools are a combination of guidelines, assessment tools, benchmarking tools, evaluations, and legislations. A brief description of each tool is presented below while Table 2 provides succinct summary of the purpose of each tool.

3.1.1 *Guidelines*

The ISEAL Alliance is a membership-based organization that has emerged as an authority on good practices for sustainability standards. **ISEAL Codes of Good Practice** are widely recognized as best-in-class guidance on how sustainability standards should be set up and managed to be credible and effective [19]. This report reviews the main ISEAL Codes of Good Practice, which are:

- **ISEAL Standard-setting Code:** This Code defines how a standard should be developed, structured, and improved over time [1].
- **ISEAL Assurance Code:** This Code presents a framework for assessing compliance with standards for the results of the assessment to be trustworthy [2].
- **ISEAL Impacts Code:** This Code provides a roadmap for robust monitoring and evaluation. It presents clear guidelines to measure progress against sustainability objectives and improve practices over time [3].

In addition to ISEAL's Code of Good Practice, this report also reviewed **ISEAL Benchmarking Criteria for Evaluating Sustainability Standards**. ISEAL's Benchmarking criteria derive from the Codes of Good Practice and offer a list of criteria to help stakeholders understand what to look for when evaluating the likely effectiveness of sustainability standards [4].

3.1.2 Assessment tools

The **Certification Assessment Tool (CAT)** was developed by World Wildlife Fund (WWF) to assess which certification schemes are best equipped to achieve conservation goals and objectives and which schemes are likely to deliver the best social and environmental impact at scale [6]. The CAT assesses a scheme's requirements, governance, rules and procedures. CAT assessments can be applied to all sectors, however, WWF has developed special criteria for terrestrial commodities and for aquaculture schemes [6].

The **Farm Sustainability Assessment (FSA)** was developed by the Sustainable Agriculture Initiative Platform (SAI Platform) to enable users to assess and measure on-farm sustainability performance in line with global best practice [7]. It was created to be adaptive to the needs of its users, therefore it could serve as a verification tool for farmers, as a benchmark, or as tool to track progress against sustainability objectives [20]. In practical terms, the FSA is a self-assessment questionnaire to assess the application of business, environmental and social sustainability farm practices [21].

The **Sustainability Assessment in Food and Agriculture (SAFA)** was developed by the Food and Agriculture Organization (FAO) to provide a common framework to assess the performance of an enterprise (be it a farm or a company), branch of a company or production site, in terms of economic, environmental, social and governance sustainability [9]. A SAFA is not a rating of product-specific sustainability, nor does it cover the use and end-of-life phases of products (e.g. at the consumer level), instead, it is a common framework to facilitate measuring progress towards the SAFA goals in a harmonized way [9]. Science-based and generic in nature, the SAFA framework is meant to be implemented at any level, national, supply chain or operational unit [9].

The **Sustainability Standards Comparison Tool (SSCT)** is an online tool developed by the German Corporation for International Cooperation (GIZ) in cooperation with ISEAL Alliance and the UN International Trade Center [10]. SSCT compares schemes and labels to show their relative strengths. It assesses certification schemes based on two themes: process and content. Process includes criteria related to the governance and management of the scheme or label, while content addresses social and environmental management criteria [22]. The tool allows the user to weigh the credibility of the schemes and labels according to the issues they care about.

3.1.3 Benchmarking tools

The **Benchmarking Platform (BP)** for Certification Schemes was developed by the EU-funded Sustainability Transition Assessment and Research of Bio-based Products (STAR-ProBio) project. The BP identifies a minimum set of requirements that any existing certification should meet as well as a set of additional requirements for improvement [12]. The BP rates schemes as having a basic, advanced or excellent performance and aims to provide a framework for continuous improvement. The BP focuses on either products derived entirely from biomass or products with minimum biomass content. The BP assesses schemes against criteria in four themes: governance and management rules, environmental, social, and techno-economic sustainability.

The **Global Benchmarking Tool (GBT)** was developed by Global Sustainable Seafood Initiative (GSSI) to recognize seafood certification schemes for the credibility and robustness of their programs [13]. The GSSI benchmarking tool aims to minimize the overall environmental impact of seafood production and to help producers, retailers and consumers understand which certifications

they can trust. The GBT assesses all certification schemes in two areas: governance and operational management [23]. Depending on the scope of the scheme, the GBT will also assess the scheme against two other areas: aquaculture and/or fisheries.

The **Sustainable Supply Chain Initiative (SSCI)** is a benchmarking tool developed by the Consumer Goods Forum (CGF) to provide clear guidance to buyers and suppliers in the consumer goods industry on which third-party auditing, monitoring and certification schemes cover key sustainability requirements and apply relevant governance and verification [14] [24]. The SSCI currently recognises independent auditing, monitoring and certification programmes that meet industry expectations on scheme management and social sustainability criteria, and will later focus on environmental sustainability [24].

3.1.4 Evaluations

The **Pusch Foundation**, in collaboration with Helvetas, the Romande Consumer Federation and WWF Switzerland, conducted an evaluation to assess the sustainability of the 31 most important food labels in the Swiss market [15]. The Pusch Foundation evaluates food labels around three themes: management, ecology and social affairs, and processes and controls. Management includes criteria for compliance with laws, environmental and social management plans, animal health management, competence, and impact analysis. Ecology and society include criteria to assess the management of environmental and social impacts (water, soil, biodiversity, climate, animal welfare, working conditions and fair-trade relations). Process and controls include criteria related to scheme governance, transparency, creation of criteria, control, certification, chain of custody, and scope and management of labels [15].

The Sustainability Monitoring and Assessment Routine (SMART) Sustainability Assessment is a tool developed by Sustainable Food Systems (SFS). SMART is adapted from the SAFA Framework to assess the sustainability of farms in a systematic manner [16]. As it is based on the SAFA framework, it measures the performance of farms in terms of economic, environmental, social and governance sustainability.

3.1.5 Legislation

The **Environmentally Preferable Purchasing Program (EPPP)** is an assessment framework developed by the United States Environmental Protection Agency (EPA) to assess the environmental performance of private sector standards and ecolabels to determine if they should be included in the EPA recommendations of specifications, standards, and ecolabels for federal purchasing [17]. The EPPP seeks to create a level playing field for manufacturers seeking to develop sustainable products by clarifying which standards and ecolabels are likely to be used by federal purchases. The EPPP assesses standards and ecolabels against criteria related to four themes: standard development, conformity, scheme management, and environmental sustainability [17].

The revised **Renewable Energy Directive (RED II)** is a legal framework for the development of renewable energy across all sectors of the EU economy [25]. The directive introduces sustainability criteria as minimum mandatory criteria for voluntary sustainability standards to ensure compliance of biofuels, bioliquids and biomass fuels with RED II. RED II sets minimum mandatory requirements in four areas: 1) sustainability and greenhouse emissions, 2) chain of custody, 3) recognition of other voluntary schemes, and 4) audit quality, scheme governance and supervision [18].

Table 2. Overview of the Purpose of the Tools Reviewed

TOOL	PURPOSE OVERVIEW
GUIDELINES	
1. ISEAL Standard-Setting Code	Transparent and accountable standards
2. ISEAL Assurance Code	Accurate assurance implementation
3. Impact Code	Measure results through monitoring and evaluation
4. Benchmarking Criteria	Consistency of benchmarking
ASSESSMENT TOOLS	
5. CAT, WWF	Assess the social and environmental impact of agriculture and forestry schemes
6. FSA, SAI Platform	Assess the application of business, environmental, and social sustainability farm practices
7. SAFA, SAFA	Assess the social and environmental impact of food and agriculture operations
8. SSCT, GIZ	Assess the credibility and sustainability of a standard organization
BENCHMARKING TOOLS	
9. BP, STAR-ProBio	Benchmark certifications schemes based on minimum governance and sustainability requirements
10. GBT, GSSI	Assess the credibility and robustness of seafood certification programs
11. SSCI, CGF	Benchmark for third-party audit or certification schemes based on sustainability best practices
EVALUATIONS	
12. ESFL, Pusch	Evaluate the performance of food labels vis-a-vis management, environmental, social and process best practices
13. SMART, SFS	Evaluate the performance of farms in terms of economic, environmental, social and governance sustainability
LEGISLATION	
14. US EPA Framework	Assess the environmental performance of ecolabels and standards for federal purchasing
15. RED II, EC	Establish mandatory sustainability requirements for VSS involved with biofuels, bioliquids, and biomass fuels.

3.2 Synthesis of evaluation criteria of shortlisted tools

This section summarizes 137 criteria which are categorized into 2 themes and 18 subthemes (see Figure 2). The criteria presented in this section reflect an aggregation of all the criteria in all the tools that were reviewed by the author. This means that not all 137 criteria are covered by each tool (see section 2.3.1 for more details on the methodology). The goal of this section is therefore to provide an in-depth overview of the type of criteria covered by tools and guidelines that were reviewed in this exercise.

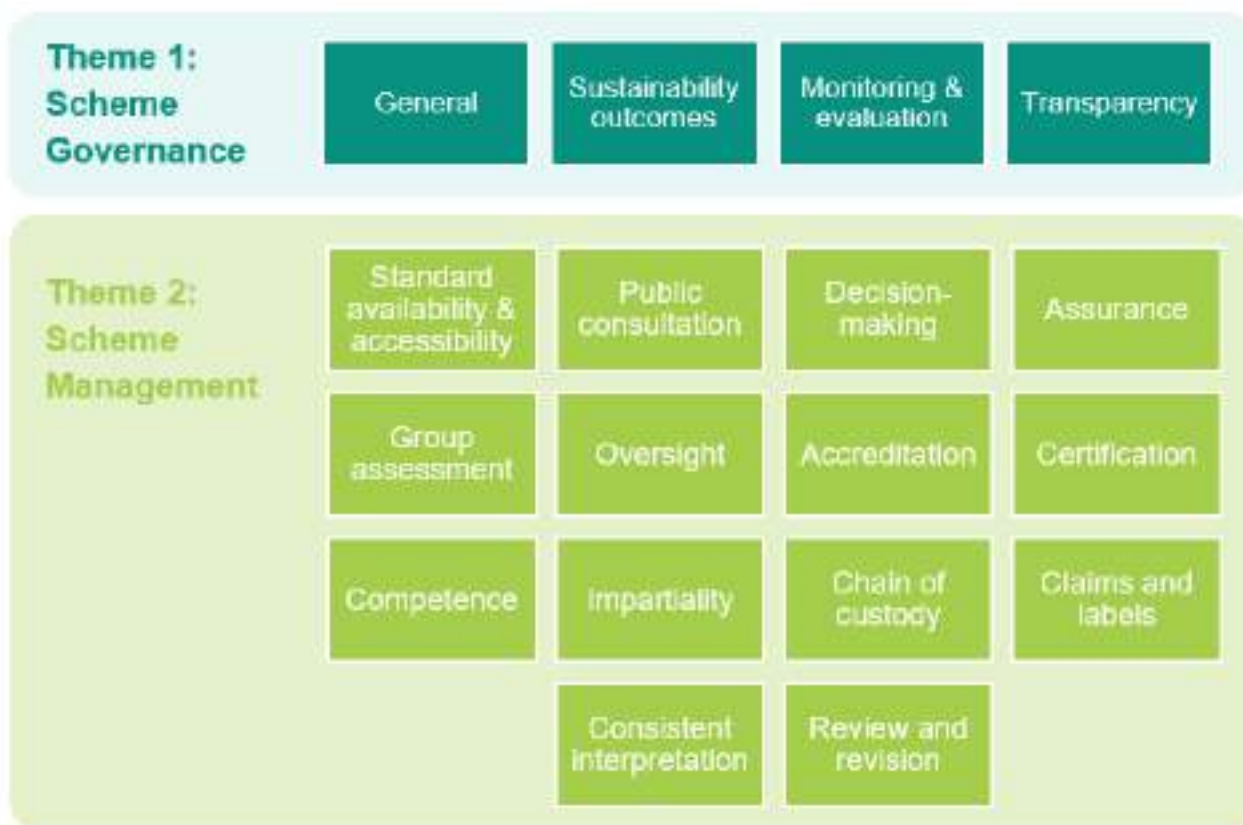


Figure 2. Overview of Themes and Subthemes

3.2.1 Theme 1: Scheme governance

General. This subtheme contains criteria related to the general administration and governance of the schemes. Criteria under this theme require scheme owners to prove legal compliance with international conventions as well as corrective action in case of any breach, and to demonstrate having and complying with a code of conduct. Criteria also require scheme owners to be ISEAL members (or at least demonstrate alignment with ISEAL's multiple guidelines), have a resourced secretariat, be registered legal entities, and have insurance and funds to cover any liabilities. Lastly, criteria under this theme require proof that the scheme does not reduce the existing rights of communities to land, water, and resources.

Sustainability outcomes. This subtheme investigates the intended sustainability goals of the schemes. Criteria require sustainability goals to be defined and based on hotspot analysis and that scheme's criteria are linked to their goals. Moreover, criteria require schemes to ask for an absolute level of performance and for requirements to ensure continued performance of the certified products, such as through the implementation of action plans that are reviewed and verified over time.

Monitoring and evaluation. This subtheme pertains to the design and implementation of a monitoring and evaluation (M&E) system. Criteria verify the existence of an M&E system and require diverse yet relevant stakeholders to participate in its development. Criteria require frequent evaluation of the performance of the schemes with explicit involvement of management and clear action plans to incorporate results from M&E data. Furthermore, criteria call for explicit documentation of intended impacts and outcomes, and for documentation of the investigation and monitoring of the unintended effects and external factors that might influence the organization's achievement of goals. Criteria require yearly and external evaluations, as well as the development

and documentation of indicators. Criteria also require the use of monitoring and evaluation results for learning. Finally, the criteria ask for an IT system to manage data to be in place.

Transparency. Transparency criteria revolve around the availability of public information regarding several areas. Criteria require that schemes make publicly available information on governance, standard-setting procedures, intended impact, the structure of the assurance system, data ownership, eligibility, application procedure, list of oversight bodies, fees, assessment methodology, policies, stakeholder involvement, consequences of non-conformity, resolved complaints, equivalence, list of certified clients, results of the assessment and of M&E data. Moreover, criteria require that evaluation results as well as any changes in scope, requirements, and rules of certification are explicitly communicated to relevant stakeholders.

3.2.2 Theme 2: Scheme management

Standard availability and eligibility. Criteria in this subtheme are centred around the availability and accessibility of the scheme's documents and processes. Criteria require standard-setting procedures to be accessible and transparent, and the terms of reference of the standard to have a clear scope, purpose, etc. Criteria verify that standards are open to all relevant operations and that they take into account special circumstances (i.e. reduced fees for smaller operations), and that the standards avoid creating market barriers. Criteria under this subtheme go into language requirements and information needed to be published for the standard to be accessible. Lastly, the criteria explicitly look for exception policies to include small or disadvantaged stakeholders/producers.

Multistakeholder participation. Criteria under this theme tackle public consultation processes and requirements. All criteria require that schemes give sufficient time and opportunity for all stakeholders to participate in the development and revision of the standards, with criteria specifying the duration of public consultations (60 days). Criteria require schemes to proactively identify stakeholders that are underrepresented or disadvantaged and ensure their participation. Criteria require that comments are taken into account and justifications are issued when comments are not accepted. Criteria also require the use of technical experts in standard development and revision.

Decision-making. Criteria in this subtheme address decision-making processes and mechanisms. Criteria require that schemes allow for a balanced representation of stakeholders in decision-making, and if there is a limit regarding who can join the decision-making body, then criteria require that the selection process is transparent and non-discriminatory and that stakeholders have the right to vote for decision-makers. Criteria also require proof that there is no dominance of any particular stakeholder in decision-making, that decisions are the result of consensus and that decisions are made public. Finally, there are criteria that require a procedure that outlines authority for assurance decisions.

Assurance. This sub-theme contains criteria related to assurance management and procedures. Criteria related to assurance management verify that schemes have specific governance produces of the assurance model and that the standard body has a risk management plan to deal with the assurance process and involved stakeholders. Criteria also require schemes to have regular management reviews of the assurance system along with processes to use data for improvement. Criteria require schemes to have a documented management system for the assurance system, and that information on the scope and durability of the certification to be available. Finally, criteria require schemes to conduct feasibility and auditability assessments.

Criteria on assurance procedures require explicit assessment methodology and procedures for sampling (or group sampling). Criteria under this subtheme also require a remediation and consequences mechanism as well as a complaints procedure. Criteria look for evidence of records

and document control including maintenance of records for at least 5 years. Finally, criteria require clear procedures for equivalence arrangements.

Oversight. Criteria in this subtheme require that schemes have in place an oversight mechanism and review procedure, proof of the independence of oversight, and clear delineation of authority over oversight bodies. Criteria also require a minimum level of competence of oversight personnel.

Group assessment. This subtheme covers criteria related to procedures and management of group certifications. Criteria under this subtheme require groups to have group assessment procedures for roles and responsibilities, entering a group, approval/removal, decision-making, code of conduct, record-keeping, assessment, complaints, and appeals. Criteria also require that the scheme have procedures to deal with groups in the guidance sections of the criteria. Finally, criteria require the existence of a group non-conformities mechanism.

Accreditation. This subtheme focuses on the accreditation process. Criteria verify that accreditation bodies are certified by the international standardisation organisation (ISO) or other relevant bodies and require that accreditation bodies make on-site audits. Criteria under this subtheme require accreditation bodies to make risk-based audits, allow and incorporate stakeholder input, and escalate minor non-conformities that are not addressed to major non-conformities. Criteria require assurance that there will be accreditation services available for everyone that needs them and that accreditation bodies undergo external audits.

Certification. This subtheme comprehensively addresses the entire certification process. Criteria verify that certification bodies have ISO certifications or comply with the ISEAL assurance code, and require that the fee structure is reasonable, transparent, and public. Criteria verify that schemes have clear conditions to suspend, withdraw, or terminate participants; mechanisms and processes to address non-compliance, complaints and appeals; and plans for corrective action. Moreover, the criteria require that the certification body is fully responsible for the actions of any subcontractors and that certification cycles should be specified and usually between 3-5 years. Criteria also require procedures for multi-site certification.

In terms of audits, criteria require a procedure to outline the frequency of audits and require that they be regular. Certification bodies are required to conduct internal, on-site, announced, unannounced, surveillance, and risk-based audits. Finally, criteria require stakeholder input into the certification process and results, and that certification bodies escalate minor non-conformities that are not addressed to major non-conformities.

Competence. This subtheme contains criteria to ensure the competence of the certification system and related service providers. Criteria require explicit personnel qualifications, ongoing training, and competence evaluation. In addition, criteria require auditor calibration and procedures for repercussions for misconduct.

Impartiality. Criteria under this subtheme verify impartiality in decision-making. Criteria require schemes to have procedures to identify and mitigate risks of impartiality or conflict of interests, and to verify auditor and assessment impartiality.

Chain of custody. This subtheme covers the identification of certified products as distinct from uncertified products and traceability throughout the value chain. Criteria verify that schemes have chain of custody verification requirements. Moreover, criteria require that all enterprises that store products are assessed and that all products are traceable from the sale of point to the customer, and that certified products are always traced. Criteria require that chain of custody records are maintained and that schemes assess the risk of origin. Lastly, criteria require information about laboratory testing methods, that the testing facility is accredited, and that there are rules on testing compliance where laboratory tests are required.

Claims and labels. Criteria under this subtheme cover the procedures and usage of claims and labels. For labels, criteria verify the existence of minimum content requirements and require that labelled products do not contain illegally harvested products. For claims, criteria verify that schemes have procedures for surveillance and monitoring of the use of claims, that claims contain enough information so they can be verified, and that legally enforceable agreements for use of claims are in place. Moreover, criteria require that claims are appropriate considering the chain of custody model.

Consistent interpretation. Criteria under this subtheme ensure consistent interpretation of standards by asking for clear language and sufficient guidance. Criteria also require that international standards be the base of national standards and that nationally applicable standards are adapted from international, regional or national principles and criteria.

Review and revision. Criteria require standards are revised and reviewed every 5 years, that stakeholders are properly involved, and that their input is fully considered. Criteria require that when there are any major revisions, all affected parties are given sufficient notice and transition period to comply.

3.3 Summary of rating and scoring methodologies

The following section summarizes the rating and scoring methodologies as well as the results visualisation choices of the shortlisted tools. The tools reviewed had varying degrees of detail regarding their rating and scoring methodologies. As a result, some of the sections below are richer than others.

3.3.1 Certification Assessment Tool (CAT)

To conduct its evaluation, the CAT organizes criteria into two themes: systems strengths and standard strengths. Systems strengths evaluate the rules and procedures that regulate how a scheme is implemented, assessed and governed. Standard strengths evaluate the requirements of a scheme. CAT considers criteria under both themes to be of equal importance therefore it gives equal weight to both themes [6]. During the evaluation, each criterion is marked as either fulfilled or unfulfilled (see Figure 3).

PART I - SCHEME OVERVIEW		
A. MISSION AND GOVERNANCE		FSC CONGO BASIN
1	The scheme is set up to continuously improve the social, environmental and economic benefits of producing a specific commodity or category of commodities. The long term social, environmental and economic impacts are defined and documented".	✓
2	Scheme core normative documents, e.g. statutes, bylaws and principles and criteria (audit manual or comparable), are publicly available on a scheme website.	✓
3	The scheme is a (full or associate) member of ISEAL.	✓
4	The scheme has a resourced Secretariat with clearly defined staff and functions related to, at a minimum; finance, standard development, communication, membership/participant handling, quality assurance and complaints and the contact details are publicly available online.	✓
5	The scheme is open to membership (and/or similar level of participation in governance) for all stakeholders who share the scheme's values and objectives.	✓

Figure 3. CAT Scoring System [26]

The author did not find a detailed manual on the evaluation methodology of the CAT, however, based on the information available the following process was assumed. Each fulfilled criterion is worth one point and each unfulfilled criterion is not given any points. When the evaluation is complete, the criteria under each subtheme are added up and divided by the total number of criteria. The score obtained under each subtheme is then normalized based on a 0-5 range. The result is then plotted

in a radar chart like the one in Figure 4 below. Figure 4 shows the theme “systems strengths” under section I: scheme overview, and the theme “standard strengths” under section II: standard.

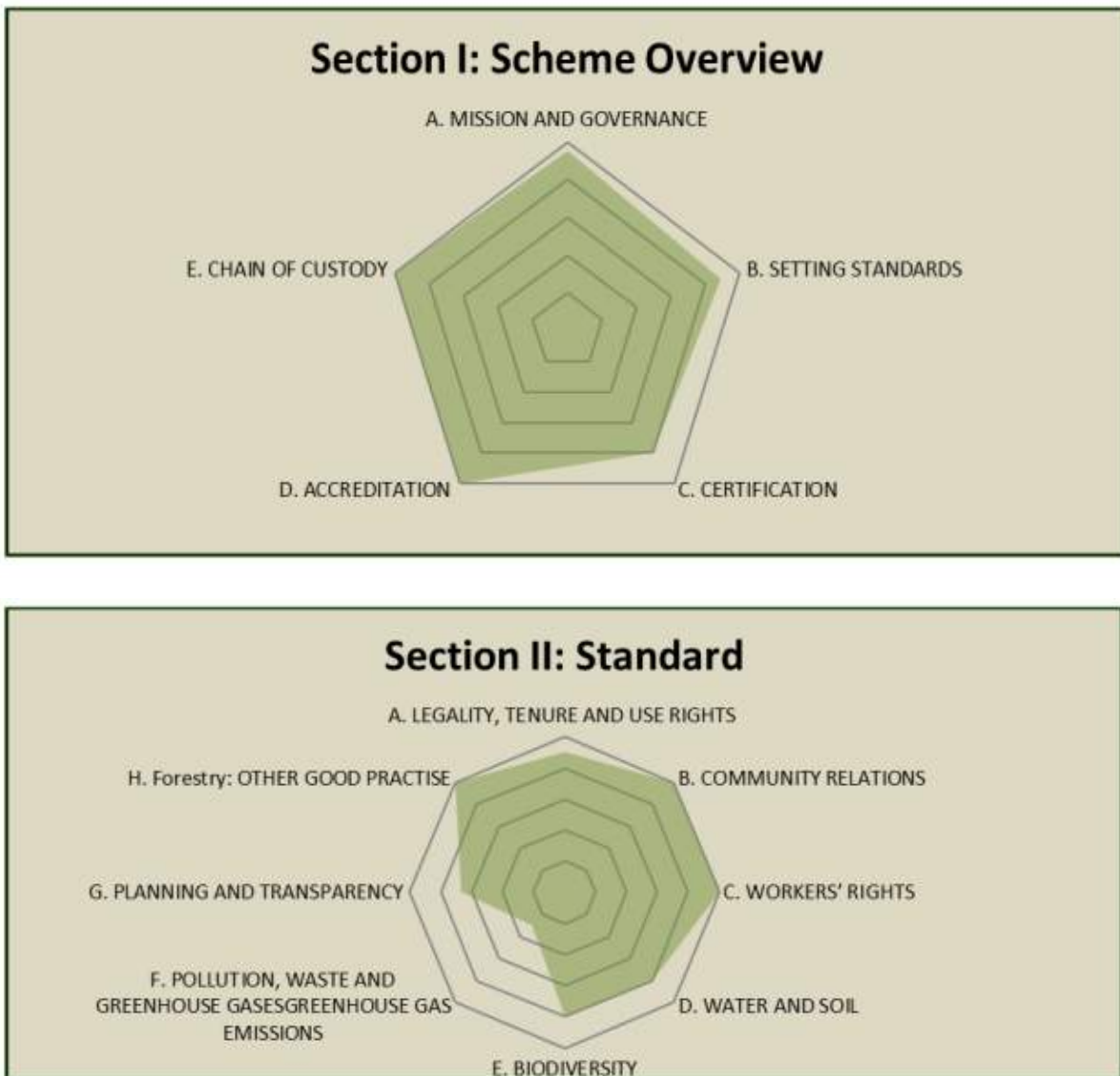


Figure 4. CAT Results Visualisation [26]

3.3.2 Farm Sustainability Assessment (FSA)

FSA is a self-assessment questionnaire (SAQ) which a farmer or farm management group can use to assess the application of sustainable farm practices. The SAQ contains yes/no questions on business, environmental and social practices [21]. Each question (and thereby each practice) are categorized as essential, intermediate or advanced (see Figure 5). Every question has a detailed description of the requirements along with guidance notes and useful links to aid the user fill the questionnaire appropriately [7].

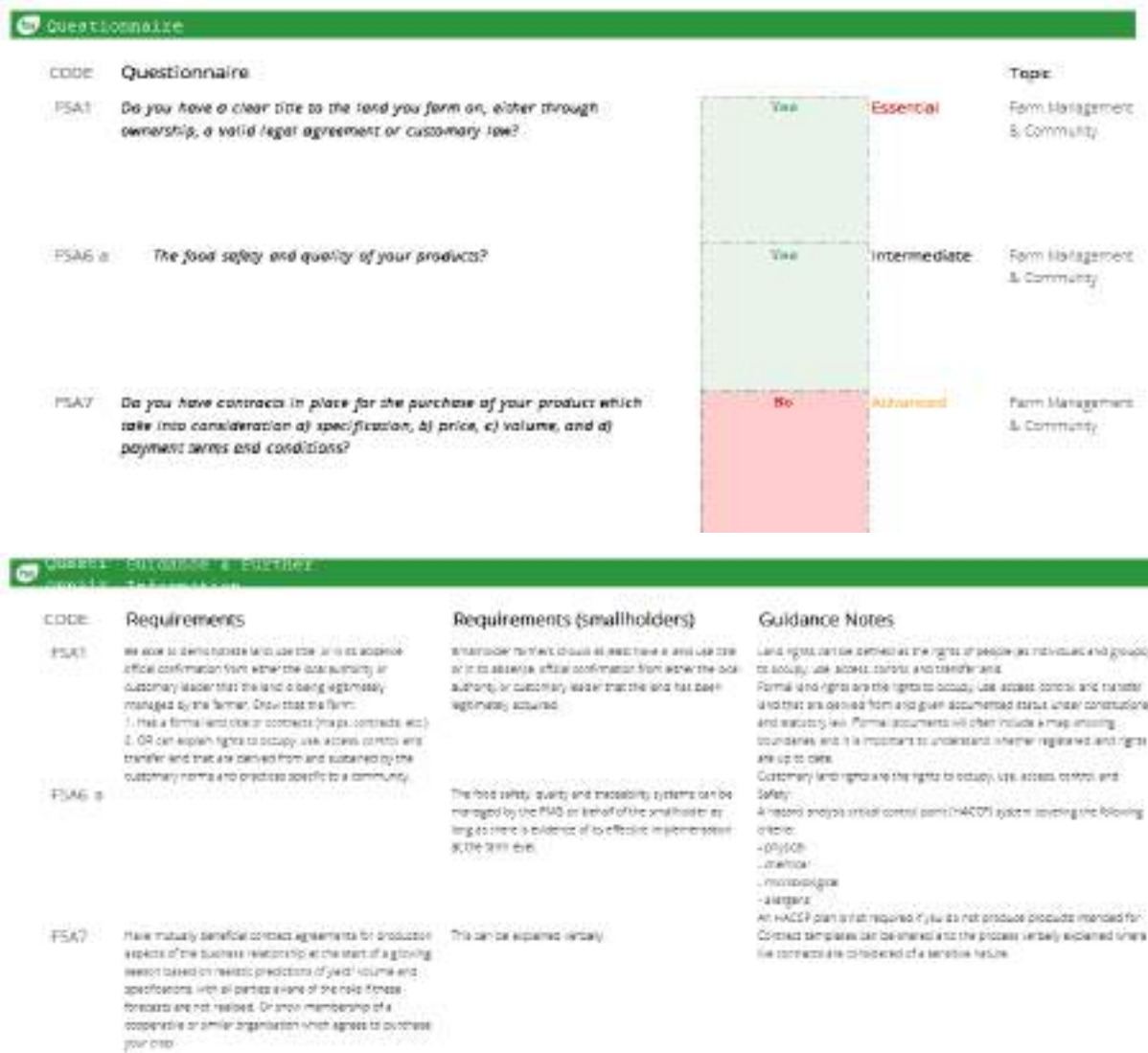


Figure 5. FSA Self-Assessment Questionnaire [7]

The Excel-based SAQ also includes a dashboard which provides an overview of the overall score as well as the score by type of question (essential, intermediate, advanced), performance claim level (bronze/silver/gold), and score per topic (see Figure 6) [7].

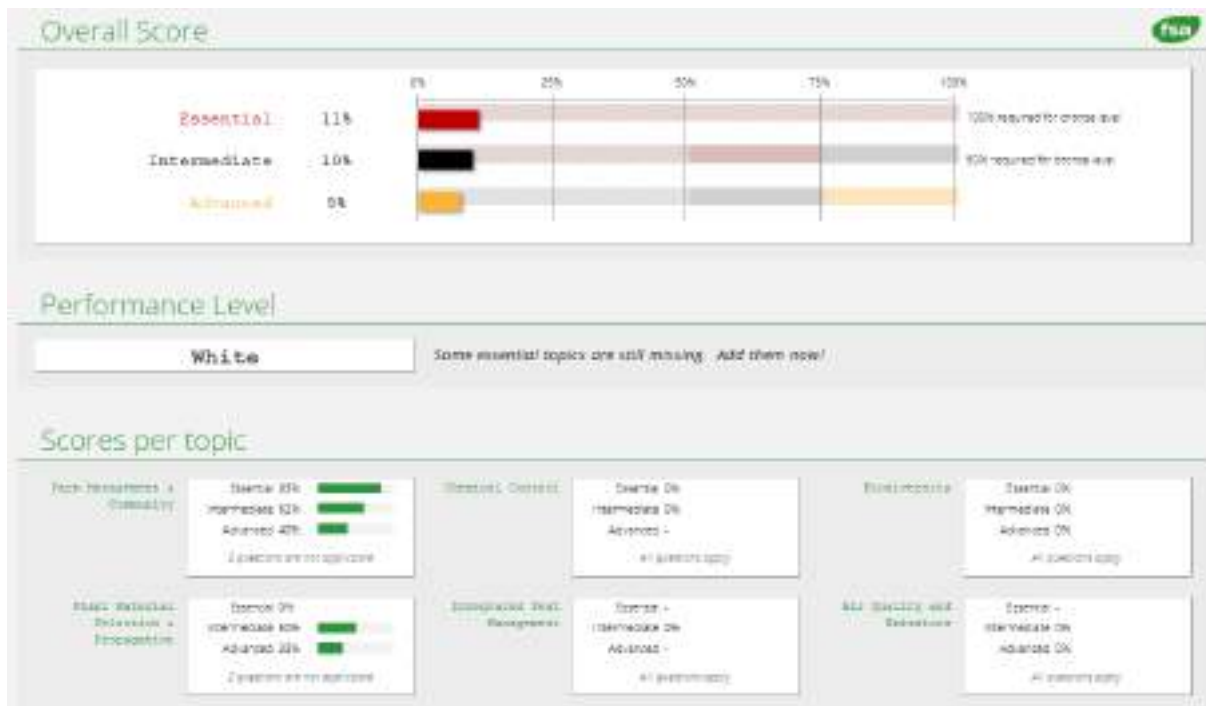


Figure 6. FSA SQA Dashboard [7]

3.3.3 Sustainability Assessment in Food and Agriculture (SAFA)

SAFA was designed as a holistic framework for the assessment of sustainability along agriculture, forestry, and fisheries value chains. It is meant to be used by a broad range of stakeholders, from small-scale farmers to multinational corporations. Therefore SAFA framework provides a detailed guiding template with 4 dimensions, 21 themes, 58 subthemes, and indicators (both default and suggested) to facilitate measuring progress towards the SAFA goals in a harmonized way (see Figure 7) [9].

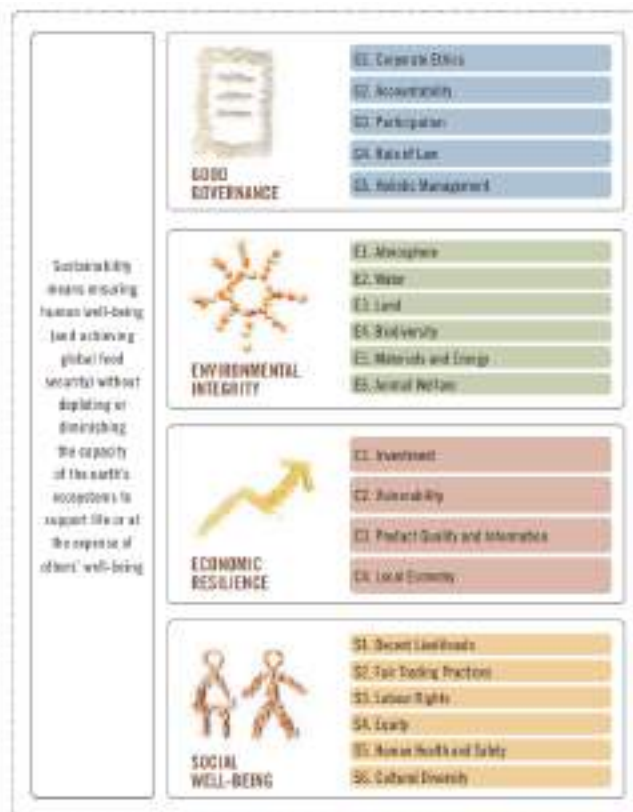


Figure 7. The SAFA Sustainability Dimensions and Themes [9]

SAFA rates each of the four sustainability dimensions on a 5-point scale from unacceptable (1) to best (5) (see Figure 8). The themes and subthemes within each dimension receive equal weighting and are rated under the same 5-point scale [9].

RATING	SCORE	PERCENTAGE SCORE
● BEST	5	80-100 percent
● GOOD	4	60-80 percent
● MODERATE	3	40-60 percent
● LIMITED	2	20-40 percent
● UNACCEPTABLE	1	0-20 percent

Figure 8. SAFA Rating Scale [9]

Although the subthemes are equally weighted, the indicators within the subthemes are weighted based on data quality. Indicators that meet high-quality criteria (which is described at lengths in the user guide) receive 3 points, indicators of moderate quality get 2 points, and low-quality indicators get 1 point. This weighted score is combined with a performance score (based on a 5-point scale, see Figure 8) to calculate a total score for each indicator [9].

Once scores have been assigned to all indicators in a sub-theme, an average score is calculated for the sub-theme based on a simple average. After all sub-themes have been scored, an overall score for each theme is calculated using the same process [9].

The result of this exercise is plotted in a radar chart (see Figure 9). The thick black line connects theme performance following a traffic light colour code: best/good (green), moderate/limited (yellow/orange) or unacceptable (red) [9].

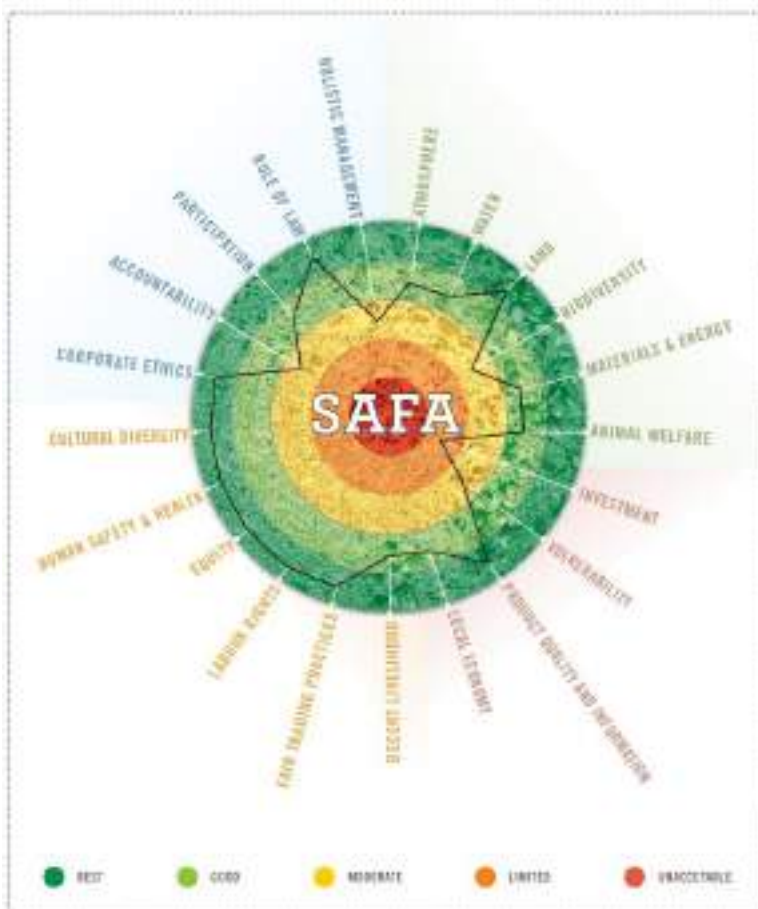


Figure 9. Visualisation of SAFA Results [9]

3.3.4 Sustainability Standards Comparison Tool (SSCT)

The SSCT organizes its criteria into three themes: credibility, environmental friendliness and social responsibility. Under each theme, criteria are further categorized as minimum or advanced criteria. SSCT uses an absolute approach to distinguish whether a criterion is fulfilled or not (yes/no answer - see Figure 10). SSCT awards the fulfilment of basic criteria with 1 point and the fulfilment of advanced criteria with 2 points [27].

Credibility		Environment		Social	
SSCT ID	Fulfilled	SSCT ID	Fulfilled	SSCT ID	Fulfilled
A.01 (MIK)	yes	E24	yes	GQ03	no
A.02 (MIK)	yes	E114 (MIK)	yes	GQ05	no
A.03	no	E30	yes	HR07	no
A.04	no	E31	yes	HR08	no
A.05 (MIK)	yes	E32	yes	HR09	no
A.06	yes	E33	yes	HR10	no

*MiK = minimum requirement

Figure 10. Example of SSCT Scoring System [27]

The SSCT is based on a start rating and the assessment process to assign stars is performed in the following steps. First, it is checked that the label fulfils the minimum requirements in the credibility theme. If the minimum requirements are fulfilled, it is checked if the criteria also fulfil all the minimum requirements in the environmental and/or social themes. The scoring of the environmental and social criteria is calculated by multiplying the degree of obligation (immediate, with transition period, recommendations) by the degree of intensity (basic, advanced) [27].

If at least 50 percent of the minimum requirements in either the social or environmental theme are fulfilled, the label receives one star. If all the minimum requirements of at least one of the two themes are fulfilled, the label receives two starts and continues to the next step [27].

At this stage, a percentage of the total achieved score for each theme is calculated (including minimum and advanced criteria). If the label fulfils more than 60 percent of all the requirements in one theme, it gets three starts for the relevant theme [27].

Finally, SSCT determines an overall rating of “good choice” versus “very good choice”. To achieve the rating of “good choice” the label needs at least 2 starts in the credibility theme and at least 2 starts in either the social or environmental theme. To achieve the rating of “very good choice” the label needs 3 starts in the credibility theme and 3 starts in either the social or environmental theme. [27] See a sample calculation sheet and overall rating in Figures 11 and 12 below.

Requirements in total	101	40	58
Number MiK	18	7	11
Fulfilled MK	18	7	11
Percentage fulfilment MiK	100%	100%	100%
Potential score (inkl. basic and advanced?)	101	46	58
Achieved score	55	27	37
Percentage fulfilment in total	54%	59%	64%
Stars	2	2	3

*MiK = minimum requirement

Figure 11. SSCT Calculation Sheet Example [27]



Figure 12. SSCT Final Evaluation Result Example [27]

3.3.5 Benchmarking Platform (BP)

The BP assesses schemes against criteria in four themes: governance and management rules, environmental, social, and techno-economic sustainability. Each theme identifies a minimum set of requirements that any existing certification should meet as well as a set of additional requirements for improvement [12]. Based on this, the BP rates schemes performance as basic, advanced or excellent. The score is a measure that indicates how the scheme performs against the other benchmarked schemes.

The classification of the criteria into one of three levels of performance and the respective scoring system is defined in case-by-case scenario by the BP in consultation with the stakeholder board. BP’s proposed performance level in the scoring system is described in Figure 13 below.

Compliance Levels		Allocation	Result / Performance
1	Basic	Clearly specified criteria & indicators/control point (non-selective!)	→ Well Implemented (=common denominator!)
2	Advanced	Clearly specified criteria & indicators/control point (non-selective!)	→ Advanced
3	Excellent	Clearly specified criteria & indicators/control point (non-selective!)	→ Excellent

Figure 13. BP Proposed Performance Level Scoring System [12]

3.3.6 Evaluation of Swiss Food Labels (ESFL)

The Pusch Foundation organized its evaluation criteria around three themes: 1) management, 2) ecology and social affairs, and 3) processes and controls (see section 3.1 for more details). To

conduct the evaluation, the evaluators developed a detailed scoring and weighting system through a standardized process that is described to have involved around 100 experts [15]. The evaluation tool assigns weights to each of the three themes, and to each group of criteria within the themes. The evaluation report does not explain the rationale behind the choice of one weight over another, however, it does explain the scoring and weighting system.

The evaluators defined each criterion as either “very important” or as “important”. Very important criteria received a maximum of four points while important criteria received a maximum of three points. The evaluators then determined that the maximum number of points that can be achieved was 200, which was distributed and weighted within the themes as shown in table X below [15].

Table 3. ESFL Maximum and Weighted Scores

Theme	Maximum theme score	Weighted score
Management	40	20%
Ecology and social affairs	100	50%
Process and control	60	30%

The criteria within the themes were also weighted. For the themes “management” and “process and controls”, all criteria were weighted equally. The management theme has four groups of criteria, each weighted at 25 percent. The process and controls theme has six criteria groups, each weighted at 16.67 percent. The criteria in the ecology and social affairs theme were weighted differently based on expert assessment of state-of-the-art research on each criterion [15]. Figure 14 below provides a snapshot of the ESFL evaluation tool developed by the Pusch Foundation.

Area of Concern	Indicator & Objective	Criteria	Category of Impact - Weighting	Criteria Score	Weighted score
C-1 Scheme governance	C-1.1 Sustainability objectives of scheme owner	The scheme owner has a sustainability-oriented mission / vision. Commitment to sustainability is being emphasized in standard with defined objectives. Pursuit of continuous improvement of ecological, social and economical gains. Members or licence holders are obligated to pursue the standard objectives.	16.7%	4.0	2.2
		Standards and scheme owner without sustainability-oriented mission		0.0	0.0
	C-1.2 ISEAL	International Scheme organisations are members of ISEAL (full or associate). National Standards comply with the ISEAL requirements (focus: relevance, measurability, result assessment)		4.0	2.2
		Criteria not addressed / not met		0.0	0.0
	C-1.3 Democratic decisions	Democratic decision making processes regarding organisational matters of standard clearly defined.		4.0	2.2
		Criteria partly met		2.0	1.1
		Criteria not addressed / not met		0.0	0.0
	C-1.4 Complaint/Appeal procedure	Complaint/Appeal procedures for standard, certification and accreditation bodies are provided, publicly described and limited in time. Confidentiality of information shall be safeguarded during the complaint process.		3.0	1.7
		Criteria partly met		1.5	0.8
		Criteria not defined		0.0	0.0

Figure 14. ESFL Scoring Template [15]

The result of the evaluation is visualized in a radial chart (see Figure 15). In the chart, the size of the slice represents the weight assigned to each theme/subtheme (highlighted by the author in red). The circumference is the maximum possible score, and the shaded area represents the score achieved by each theme/subtheme. The dotted line indicates 50 percent.

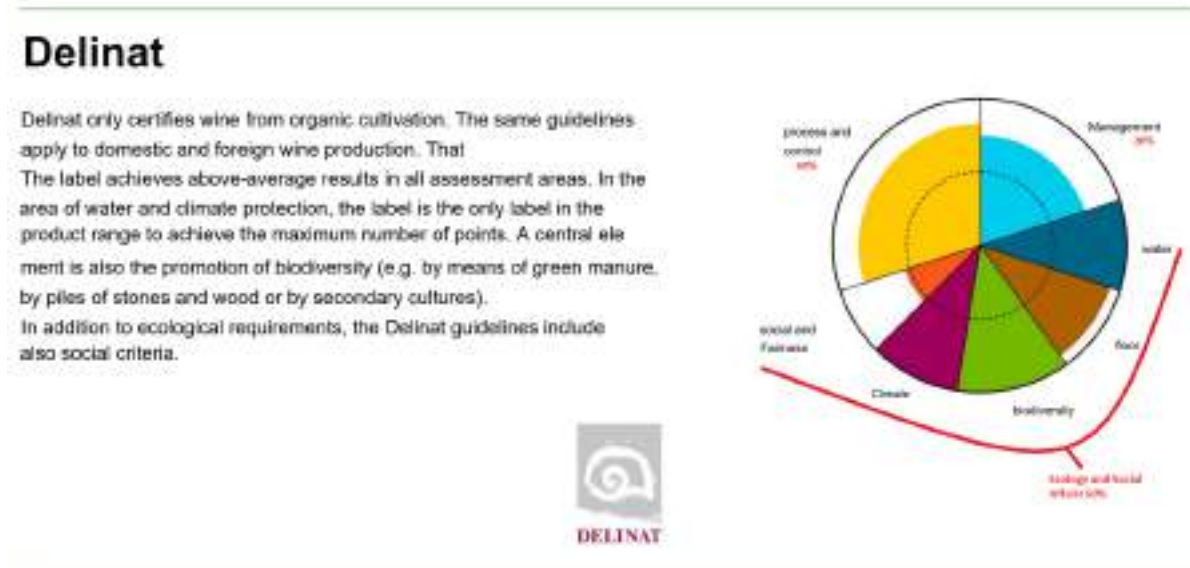


Figure 15. ESFL Results Visualisation [15]

3.3.7 Sustainability Monitoring and Assessment Routine (SMART)

The Sustainability Monitoring and Assessment Routine (SMART) Sustainability Assessment is a tool adapted from the SAFA Framework to assesses the sustainability of farms in a systematic manner [16]. It follows the same rating and scoring procedure as described under the SAFA subsection therefore it will not be repeated in here.

The SMART tool is an interesting case study because in addition to plotting its results in a radar chart as suggested by SAFA (see Figure 16), it produces a summary sheet which justifies the rating for each theme by highlighting aspects that had either a positive or negative impact on the rating (see Figure 17).



Figure 16. SMART Results Visualisation [16]

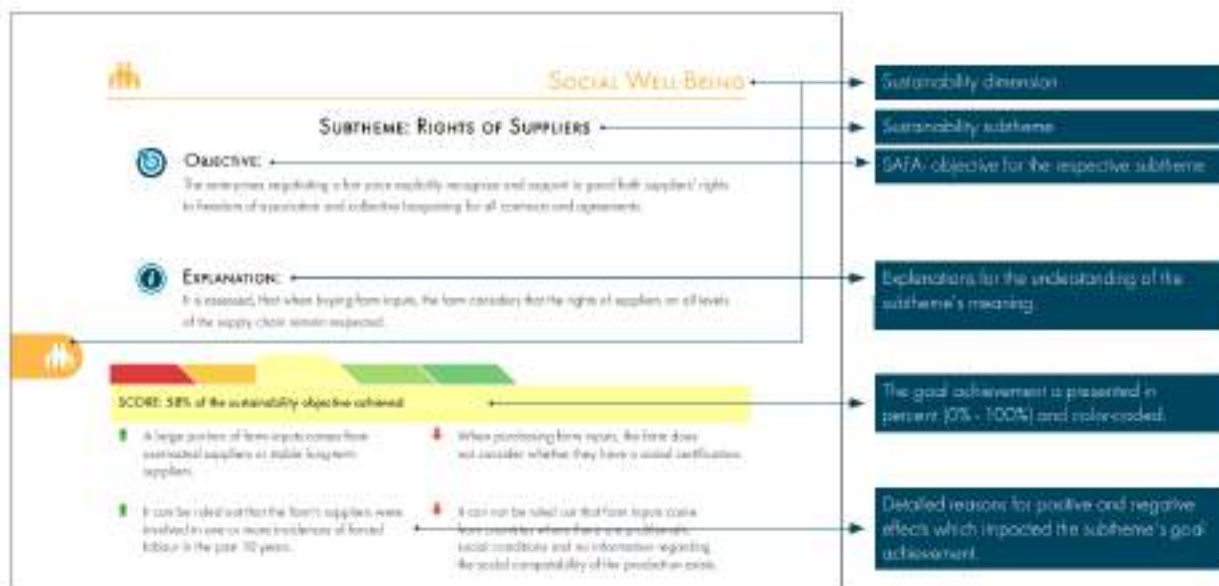


Figure 17. SMART Summary Sheet [16]

4. Discussion and conclusions

The goal of this review was to identify existing criteria, rating and scoring methodologies and visualisation approaches that could be leveraged by the project team when developing its monitoring system. The authors reviewed the criteria of 13 tools, and the rating and scoring methodologies and visualisation of results of 7 tools.

The monitoring system that will be developed under task 3.2 is meant to assess the effectiveness and robustness of sustainability certifications and labels, as well as define minimum requirements. The SUSTCERT4BIOBASED team is in the process of defining effectiveness and robustness and determining the scope and aim of the monitoring system. This task 3.1 contributes to our thinking on how to monitor effectiveness and robustness while task 1.2 contributes to having an overview of sustainability principles and criteria.

At this point, the team understands effectiveness as the extent to which the scheme/label achieves, or is expected to achieve, its objectives and bring results. Effectiveness answers the question: Is the scheme or label achieving its objectives? On the other hand, robustness is the extent to which the assurance system in place is implemented credibly and well. It answers the question: Does the scheme/label deliver accurate assessment of compliance (not prone to misinterpretation or fraud)?

The majority of the criteria that were reviewed in this exercise focus on the process to develop, manage, maintain, and revise a certification scheme or label. This type of criteria would be adequate to assess the robustness of a scheme based on the definition discussed above. The review also highlighted criteria that can be leveraged to assess the achievement of results (effectiveness), although to a much lesser extent than for robustness, and primarily in the subthemes of sustainability outcomes and monitoring and evaluation. Figure 18 in the following page illustrates which subthemes (and therefore criteria) can be leveraged to assess effectiveness and robustness under our monitoring system. The exact criteria that will be used as input for our monitoring team was not determined during this exercise.

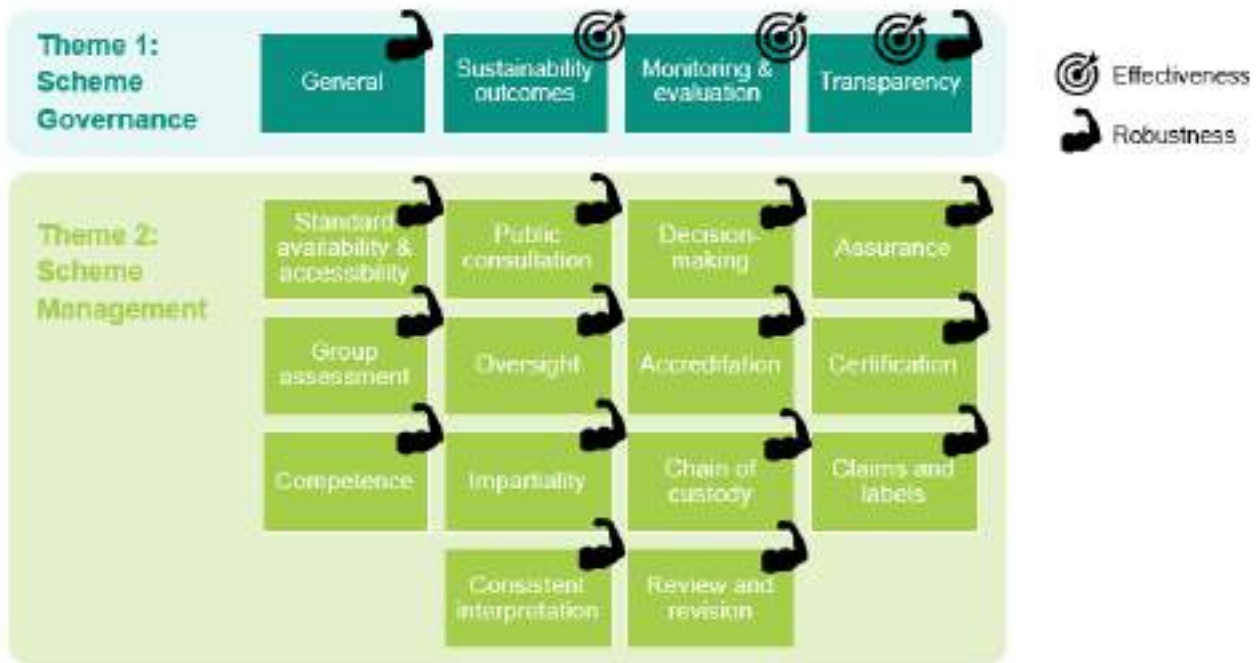


Figure 18, Initial classification of the how the criteria reviewed could be leveraged to measure effectiveness or robustness under our monitoring system

Most tools were found to differentiate criteria in at least two categories, with the most common classification being basic versus advanced criteria. Creating different tiers of criteria can serve to establish minimum performance requirements while still creating a roadmap for improvement, therefore is an approach that will be considered by our monitoring system.

When it comes to rating and scoring, the methodologies across tools varied in two main ways. First, some tools have an absolute approach to determine whether criteria are being met (yes/no) while others rate criteria in a more nuanced way (met/met partially met, etc). Second, some tools assign different weights to different criteria and/or themes while others didn't. The initial thinking of the team is that on a scale where 1 is absolute ratings and equally distributed weights, and 10 is ratings of 5 point-scales and weights differ from the criteria all the way to the theme level, our monitoring system will likely be a 6 or 7.

In other words, we think there is a lot of value in providing a more nuanced rating and scoring system, not only because it will yield richer information and thus higher usability, but because there are themes that will be of higher importance than others and data points that will be of better quality than others. Therefore, differentiation will be important to conduct a valid and credible assessment. However, there is a risk to create a system that is too complicated to implement in practice. Therefore, the team will actively weigh in the trade-offs and seek to develop a monitoring system that provide useful results while being feasible to implement.

Lastly, the visualization aspect of the monitoring system will be important. The team will aim to create two types of visuals: an overview visual like the FSA dashboard, and an overview sheet like the SMART project. An overview visual will provide the user with a synopsis of the assessment results, while the overview sheet will add high level details to help the user understand the results.

In conclusion this exercise was extremely helpful in deepening our understanding of the existing tools that assess certification schemes and labels. This review provided the team with a strong base of criteria that will be leveraged to develop our monitoring system as well as with powerful ideas for the rating, scoring and visualisation of the results of our to-be monitoring system.

5. References

- [1] ISEAL, 'Setting Social and Environmental Standards Version 6.0 (Standards Setting Code)'. 2014. [Online]. Available: https://www.isealalliance.org/sites/default/files/resource/2017-11/ISEAL_Standard_Setting_Code_v6_Dec_2014.pdf
- [2] ISEAL, 'Assuring Compliance with Social and Environmental Standards Version 2.0 (Assurance Code)'. 2018. [Online]. Available: https://www.isealalliance.org/sites/default/files/resource/2018-02/ISEAL_Assurance_Code_Version_2.0.pdf
- [3] ISEAL, 'Assessing the Impacts of Social and Environmental Standards Systems Version 2.0 (Impacts Code)'. 2014. [Online]. Available: https://www.isealalliance.org/sites/default/files/resource/2019-06/ISEAL_Impacts_Code_Version_2.0.pdf
- [4] ISEAL, 'Assessing the Impacts of Social and Environmental Standards Systems Version 2.0 (Benchmarking Guide)'. 2014. [Online]. Available: https://www.isealalliance.org/sites/default/files/resource/2019-06/ISEAL_Impacts_Code_Version_2.0.pdf
- [5] J. Zeuthen and L. Ludvigsen, *Global Third-Party Certification and Labelling*, First edition. LETZ CERTI, 2021.
- [6] World Wildlife Fund, 'Certification Assessment Tool Version 4 Factsheet'. 2016. [Online]. Available: https://wwf.panda.org/wwf_news/?228430/WWF-Certification-Assessment-Tool-CAT
- [7] Farm Sustainability Assessment, 'Self-assessment Questionnaire Version 3.0'. Apr. 2021. [Online]. Available: <https://saipatform.org/resource-centre/fsa/>
- [8] MSI Integrity and International Human Rights Clinic at Harvard Law School, 'MSI Evaluation Tool, V.1.0'. 2017. [Online]. Available: https://www.msi-integrity.org/wp-content/uploads/2017/11/MSI_Evaluation_Tool_2017.pdf
- [9] Food and Agriculture Organization, 'Sustainability Assessment of Food and Agriculture Systems Guidelines Version 3.0'. 2014. [Online]. Available: <https://www.fao.org/3/i3957e/i3957e.pdf>
- [10] GIZ, 'Sustainability Standards Comparison Tool', 2021. <https://www.ssct-revision.org/nachhaltigkeit/de/home/file/fileId/389/name/Factsheet%20SSCT%20EN.pdf>
- [11] United Nations Conference on Trade and Development, 'Voluntary Sustainability Standards Assessment Toolkit'. 2020. [Online]. Available: https://unctad.org/system/files/official-document/ditctabinf2020d4_en.pdf
- [12] STAR ProBio, 'STAR ProBio Deliverable 8.2 Blueprint of sustainability certification schemes for bio-based products'. 2020. [Online]. Available: http://www.star-probio.eu/wp-content/uploads/2017/04/D8.2_SAT-ProBio-blueprint_final-report_3-scalone.pdf
- [13] Global Sustainable Seafood Initiative, 'Global Benchmark Framework version 2.0'. NA. [Online]. Available: <https://www.ourgssi.org/benchmarking/>
- [14] The Consumer Goods Forum, 'SSCI Benchmarking Requirements Version 1.1 Part II - AI: Processing/Manufacturing Scope (Requirements for the Management of Schemes)'. NA. [Online]. Available: https://www.theconsumergoodsforum.com/wp-content/uploads/2019/10/SSCI_Part-II_AI-Processing-and-Manufacturing_Scheme-Management_fin_v1.1.pdf
- [15] Pusch Foundation, 'Evaluation des labels alimentaires', *Labelinfo.ch*, NA. <https://www.labelinfo.ch/fr/projets-devaluation/-evaluation-des-labels-alimentaires>
- [16] Sustainable Food Systems, 'Sustainability Monitoring and Assessment RouTine (SMART) Sustainability Assessment'. 2015. [Online]. Available: https://www.fibl.org/fileadmin/documents/de/themen/nachhaltigkeitsanalyse/smart/20170819_S_MART_ExampleReport_EN_MedResolution.pdf

- [17] U.S. Environmental Protection Agency, 'EPA Environmentally Preferable Purchasing Program: Framework for the Assessment of Environmental Performance Standards and Ecolabels for Federal Purchasing. Version 2.0'. 2022. [Online]. Available: https://www.epa.gov/system/files/documents/2022-02/updated-framework_020222.pdf
- [18] European Commission, 'Commission Implementing Regulation (EU) on Regulation on Rules to Verify Sustainability and Greenhouse Gas Emissions Savings Criteria and Low Indirect Land-use Change-Risk Criteria'. 2021. [Online]. Available: [https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=PI_COM:Ares\(2021\)4234307](https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=PI_COM:Ares(2021)4234307)
- [19] 'ISEAL Codes of Good Practice', ISEAL. <https://www.isealliance.org/defining-credible-practice/iseal-codes-good-practice>
- [20] Farm Sustainability Assessment, 'User Guide version 3.0'. 2022. [Online]. Available: https://saipatform.org/wp-content/uploads/documents/17164/fsa-user-guide_final.pdf
- [21] Farm Sustainability Assessment, 'Implementation Framework Version 3.0'. Apr. 2021. [Online]. Available: https://www.globalgap.org/.content/.galleries/documents/210426_FSA_Implementation_Framework_V3.0_en.pdf
- [22] K. Komives, 'Sustainability Standards Comparison Tool, ISEAL Alliance', NA. [Online]. Available: <https://unfss.org/wp-content/uploads/2013/09/ssct-presentation-for-ssi-meeting.pdf>
- [23] Global Sustainable Seafood Initiative, 'Global Benchmarking Tool: GSSI Benchmark Procedures Version 2.0'. 2022. [Online]. Available: <https://www.ourgssi.org/wp-content/uploads/2022/07/GSSI-Benchmark-Procedures-2022.pdf>
- [24] The Consumer Goods Forum, 'The Sustainable Supply Chain Initiative', *About*. <https://www.theconsumergoodsforum.com/social-sustainability/sustainable-supply-chain-initiative/about/>
- [25] European Commission, 'Renewable energy directive', *European Commission's Website*. https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en
- [26] World Wildlife Fund, 'Assessment of Forest Stewardship Council regional Congo Basin certification scheme'. 2018. [Online]. Available: https://wwf.panda.org/wwf_news/?340111/WWF-assessment-of-FSC-regional-Congo-Basin-and-PAFC-Gabon-certification-schemes
- [27] Siegelklarheit, 'Sustainability Standards Comparison Tools (SSCT) calculation example for the overall assessment'. 2021. [Online]. Available: https://www.siegelklarheit.de/static/media/SGK_Berechnungsbeispiel_EN_RK%20NEU.pdf

Annex A: Criteria Synthesis Matrix

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
General	ISEAL membership		x						x		
	Compliance with ISEAL Code of Good Practice		x								
	Existence of resourced secretariat		x								
	Commitment to code of conduct		x		x				x	x	
	Legal compliance with international conventions		x	x					x		
	Corrective actions in case of breach of legal compliance			x							
	Scheme does not reduce existing rights of communities			x							
	Liability and financing	x				x	x	x			
	Public info on financing structure	x			x				x		x
	Oversight bodies assurance providers are legally incorporated	x		x	x	x	x	x		x	
Sustainability outcomes	Stakeholder input on intended impact	x									x
	Clearly defined sustainability goals	x	x	x	x	x	x		x		
	Criteria linked to goals	x			x	x	x		x		
	Claims are consistent with goals	x			x	x	x		x		
	Criteria requires an absolute level of performance		x						x	x	
Scheme requires improved performance				x				x			
Monitoring and evaluation	Working monitoring and evaluation (M&E) system	x	x		x	x	x		x	x	
	Clear scope and boundaries of the M&E system	x			x	x	x				
	Data Management and IT system in place	x			x			x	x		
	Stakeholder consultation to provide input to M&E system	x				x	x				

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	Assessment of intended impact and outcomes	X	X	X	X	X	X		X		X
	Development of a theory of change	X			X						X
	Understanding unintended effects	X		X					X		X
	Definition of influencing factors	X		X							X
	Performance monitoring as well as outcome & impact evaluations	X		X		X	X		X	X	
	Definition and documentation of indicators	X	X			X	X			X	
	Performance monitoring and reports (ongoing data collection)	X				X	X		X		
	Third party evaluations	X	X								
	Internalising learning and improving	X	X	X	X						
	Responding to and Using M&E Results (management response)	X	X	X	X					X	
Transparency	Public Information on scheme's desired impacts	X	X		X	X	X		X		
	Publicly available information on: the structure of the assurance system, data ownership, eligibility, application procedure, list of oversight bodies, assessment methodology, policies, stakeholder involvement, consequences of non-conformity, resolved complaints, equivalence, list of certified clients, results of the assessment	X	X	X	X	X	X	X	X	X	X
	Transparent standard-setting procedures	X	X			X	X	X	X	X	
	Communication of evaluation results to relevant stakeholders	X		X							
	Transparency of evaluation results	X	X		X					X	
Standard availability and accessibility	Schemes are accessible for free to all	X			X	X	X			X	
	Access to printed versions	X									
	Info required to ensure scheme is accessible	X				X	X				

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	Language requirements	x				x	x				
	Scheme is open to all types of operations		x			x	x				x
	Scheme takes into account special circumstances	x	x			x	x				x
	Scheme avoids creating market barriers					x	x				x
	Scheme has exception policies	x	x							x	
	Scheme clearly outlines eligibility requirements	x	x		x	x	x		x		x
Multi- stakeholder Participation	Stakeholder Identification	x	x	x	x	x	x	x	x		x
	Time of opportunity to participate in public consultation (PC)	x	x	x	x	x	x	x	x		x
	PC is open to all stakeholders	x	x	x	x	x	x	x	x		x
	Underrepresented and disadvantaged groups are proactively engaged in PC	x	x	x	x	x	x		x		x
	Scheme takes PC comments into Account	x		x	x	x	x		x		x
	Scheme respond to PC comments	x		x		x	x	x	x		x
	Use of technical experts in PC	x	x			x	x	x	x	x	
Decision- making	Stakeholder balance in decision making	x	x		x	x	x		x	x	x
	Application of decisions is transparent and non-discriminatory	x	x		x	x	x				
	Decision making is driven by consensus	x	x		x	x	x		x	x	x
	Lack of dominance of any one interest in decision making	x	x		x	x	x		x		x
	Decision making mechanism and procedure	x				x	x		x		
Assurance	Accountability (governance of assurance system procedures)	x			x	x	x				
	Operating procedures (Defined scope and durability info in certification)	x			x			x			
	Operating procedures (Assessment methodology)	x			x	x	x	x			

Theme	Criteria	ISEAL Impact Assurance Standard-Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	Operating procedures (Sampling within assessment (w or w/o groups))	X				X	X	X			
	Complaints & appeals (procedure)	X	X	X	X	X	X	X		X	X
	Remediation and consequences (mechanism)	X	X	X		X	X		X	X	
	Operating procedures (Documented management system)	X			X	X	X	X		X	X
	Records and document control	X	X			X	X	X		X	X
	Records (keep records on files for 5 years)	X				X	X			X	
	Feasibility Assessment (auditability)	X			X	X	X				
	Risk management plan	X		X		X	X		X		X
	Performance management (Management reviews of assurance system)	X		X		X	X				
	Performance management (Improvement feedback loop)	X		X		X	X				
	Equivalence arrangements	X			X						X
Oversight	Oversight mechanism & review procedure	X			X					X	
	Independence of oversight	X			X				X		
	Competence of oversight personnel	X			X						
	Authority of oversight bodies	X			X					X	
Group Assessment	Group assessment procedures for R&R, entering group, approval/removal, decision-making, CoC, record keeping, assessment, complaints, appeals	X			X					X	
	Shared management system				X						
	Group non-conformities	X							X	X	
Accreditation	ISO certification	X	X			X	X	X		X	X
	Availability of accreditation services					X	X		X		
	External audits					X	X				

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	Onsite audit				X	X	X	X			
	Risk-based audits		X								
	Stakeholder input		X								
	Minor vs major non-compliance		X								
Certification	ISO certification or ISEAL assurance code		X		X	X	X	X	X	X	X
	Fee structure				X	X	X		X		X
	Responsibility for outsourcing	X	X			X	X	X		X	
	Conditions to suspend/withdraw/terminate		X		X	X	X	X	X	X	X
	Certification cycle 3-5 years		X			X	X	X	X	X	
	Multisite certification					X	X	X		X	
	Surveillance & monitoring		X			X	X	X	X	X	X
	Non-conformities procedures and mechanism	X				X	X		X		
	Certification non-compliance		X		X	X	X	X		X	X
	Minor vs major non-compliance		X							X	
	Corrective action		X		X	X	X	X	X	X	X
	Mechanism for complaints and appeals	X	X	X	X	X	X	X	X	X	X
	Resolving complaints	X	X	X	X	X	X	X	X	X	X
	Complaints decision disclosure	X	X	X	X	X	X	X	X	X	X
	Frequency of audits	X	X		X	X	X		X	X	
	Risk-based audits, intensity		X		X			X		X	
	Internal audits	X				X	X				
	Full audits include office visits	X	X	X	X	X	X	X			
Audits stakeholder input	X	X	X	X	X	X		X			
Audits unannounced audits		X		X			X	X			
Competence	Personnel qualifications	X			X	X	X	X	X	X	X
	Building competence	X			X	X	X	X		X	X

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	Auditor calibration	X				X	X	X			
	Evaluation of competency	X			X	X	X	X		X	
	Repercussions for misconduct	X			X						X
Impartiality	Risks of impartiality and conflict of interest	X			X	X	X		X	X	
	Auditor impartiality	X			X				X	X	
	Impartiality in the assessment	X			X				X	X	
	Decision-making	X			X	X	X			X	X
Chain of Custody	All certified products identified at all times		X		X	X	X		X	X	
	Chain of Custody verification requirements	X	X		X	X	X		X		X
	Assessment of enterprises that store products	X			X	X			X		X
	Traceability of products	X			X	X				X	X
	Maintenance of records				X	X	X			X	
	Assessing risk of origin		X			X				X	
	Lab testing requirement				X						
Claims and Labels	Requirements for use of claims and labels, including min content	X	X		X		X	X	X	X	
	Legal agreement for use of claims	X			X		X	X			
	Claims and chain of custody model	X			X						
	Claims can be verified	X			X		X				
	Surveillance to monitor use of claims	X	X		X		X	X			
	Products do not contain illegally harvested or traded		X								
Consistent Interpretation	Clear language	X					X				
	Sufficient guidance	X		X			X	X			
	International standards as base for national standards	X	X						X		

Theme	Criteria	ISEAL Impact Assurance Standard- Setting Benchmarking	CAT WWF	SAFA FAO	SSCT GIZ	BP STAR	GBT GSSI	SSCI CGF	ESFL PUSCH	RED EC	EPPP EPA
	National standards are consistency with international standards	x	x						x		
	Local Applicability - interpretative guidance to take local conditions to account	x	x	x			x		x		
	Local Applicability - recognition of existing standards	x	x							x	
Review and revision	Review period	x	x	x			x	x	x		
	Publish date of review	x					x				
	Process to submit comments for revisions/clarifications	x					x				
	Consideration and response to input	x	x				x	x			
	Transition Period (adequate time to comply)	x			x		x				



About SUSTCERT4BIOBASED

SUSTCERT4BIOBASED is an EU funded (Horizon Europe) project aiming at defining and promoting the adoption of effective and robust sustainability certification schemes and business-to-business labels for industrial biobased systems to support tracing the sustainability (environmental, social, economic) of biobased products along the value chains and trades within the EU and globally for responsible production and consumption. This objective is realised by the development of a monitoring system, mapping of the current situation in global trade flows of biological resources and biobased products, and feasibility assessment from the adoption of certification schemes and labels considering actual economic as well as internalized environmental and social costs and benefits. The results of the project are leveraged to provide recommendations to four key target groups: policy makers, sustainability system community, industrial biobased value chain actors, and regional bioeconomy stakeholders. These ambitions are addressed by a strong, well-balanced and multi-disciplinary consortium comprised of 5 complementary partners. SUSTCERT4BIOBASED thereby supports the development of harmonized system requirements, continuous improvement of sustainability certification schemes and labels and contributes towards establishing a circular, climate-neutral and sustainable biobased industry.

PARTNERS



Stichting Wageningen Research (WR)
www.wur.nl



Fundacion Circe Centro de Investigacion de Recursos y Consumos Energeticos (CIRCE)
www.fcirce.es



White Research SRL (WHITE)
white-research.eu



Environmental Coalition on Standards (ECOS)
www.ecostandard.org



Control Union Certifications Germany GmbH (CU)
www.controlunion-germany.com

CONTACT US

info@sustcert4biobased.eu

- Sustcert4biobased
- @SUSTCERT4BIO
- SUSTCERT4BIOBASED
- SUSTCERT4BIOBASED

VISIT

www.sustcert4biobased.eu