Utrecht University

Master's Thesis:
The Quality of Standards

An Assessment Framework to Measure the Quality of Organisations Producing and Maintaining Voluntary Sustainability Standard Systems

Justin Klinge (5793599)

Supervisor
Dr. Walter Vermeulen

2nd Reader
Dr. Agni Kalfagianni

27th of May 2018

Master's Thesis Proposal (30 ECTS)
MSc Sustainability Development - Earth System Governance
By Justin Klinge (5793599)
justinklinge93@gmail.com / j.klinge@students.uu.nl
Windmühlenesch 15
48565 Steinfurt
Germany

Word Count: 20411
(excl. Title Page, References and Appendix)

Abstract

This research paper has tried to create an assessment framework for good governance in voluntary sustainability standard systems. After introducing standard systems and the good governance principle as well as the development and history thereof, the paper moves on to describe the usefulness of a combination of these two elements. Having identified a dominance of good governance assessments from the public and private sector, the research proceeded to analyse those existing initiatives, effectively extracting individual criteria that seemed, based on a review of the functioning of standards, suitable to the standards context. With a collection of twenty-seven public and private initiatives and researches with an uncountable number of criteria, the process of creating the actual framework was initiated, eventually leading to a set of 38 criteria that were split into five categories created and named after the five good governance principles of the UNDP; Legitimacy and voice, direction, performance, accountability and transparency and fairness. Finally, to test how accurate this first version of the framework was, it was tested on the FSC, PEFC, UTZ/RA and GCP all of which are large standards operating in either the forestry or coffee sector. Despite a planned two-round assessment phase of which the first would be based on publicly available information and the latter based on expert input from those organisations, it was only possible to do the first round of this assessment due to time constraints of the research. In the end, the assessment framework that has been created before the assessment phase is evaluated and, although it is quite incomplete at this point, should be able to provide an objective good governance assessment for any type of voluntary sustainability standard. Once more input from experts and the conclusions of this paper are incorporated, a more complete version of the framework could be produced and made universally applicable as an objective tool to evaluate good governance in voluntary sustainability standard systems.

Glossary

Accreditor
- Provide accreditation to auditors
- Often employed by standard setting organisations to improve and control the activities of their auditors
- May be a third party but can also be part of the standard setting organisations

Auditor
- Typically, an independent organisation that undertakes the certification process of producers based on the standards of a standard setting organisation
- Usually seeks accreditation and thereby adopts certain procedures and/or guidelines for the certification process
- May be used interchangeably with “certification body” in the document

Good governance
- A set of context specific dimensions, its fulfilment used to evaluate the governance ability of an entity

Standard setting organisations
- A set of/an organisation(s) in charge of managing the standard initiative, code of conduct or audit protocol
- In a single system may be multiple standard setting organisations who will then have a highest/international organisation that takes care of decision-making and related day-to-day activities
- May be used interchangeably with “standard setter” or referred to as “organisation” in this paper

Voluntary sustainability standard
- A set of criteria and indicators usually for a certain (type of) product, indicating an increased quality of often intangible product attributes
- May be used interchangeably with “standard”, “voluntary standard”, “sustainability standard”, “standard system” in the document

Abbreviations

FSC	Forest Stewardship Council
GCP	Global Coffee Platform
NGO	Non-governmental Organisation
PEFC	Programme for the Endorsement of Forest Certification
RA	Rainforest Alliance
UN	United Nations
UNDP	United Nations Development Programme

Table of Contents

Figures and Tables

1. Introduction

1.1 Background

Voluntary sustainability standards are a governance tool that has risen to popularity throughout the past two decades. The United Nations Forum on Sustainability Standards refers to the World Trade Organisation's Agreement on Technical Barriers to Trade, defining an (international) standard as; "Document approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for products or related processes and production methods, with which compliance is not mandatory. It may also include or deal exclusively with terminology, symbols, packaging, marking or labelling requirements as they apply to a product, process or production method" (UNFSS, 2013, p. 14). However, it must be noted that this definition is not one for voluntary sustainability standards specifically and therefore does not entail sustainability. Recognition of voluntary sustainability standards as international standards would greatly improve its standing, since voluntary sustainability standards sometimes are seen as technical barriers to trade, while international standards have been acknowledged as being favourable to free and international trade, with governments being encouraged to make use of such standards (UNFSS, 2013).
Komives and Jackson (2014, p. 3) complement the definition of a standard by elaborating that voluntary sustainability standards are "non-government initiatives that seek to drive sustainable production and consumption by creating market demand for sustainable products, and a supply to meet that demand. They help buyers (both consumers and businesses) identify sustainably-produced products, and they guide producers [...] and others in the choice of sustainable practices." Whether or not voluntary sustainability standards lead to sustainability can be debated and is not an uncontested statement in literature, however, it is their intention to support sustainable development efforts.

1.1.1 History and Development of Standards

Originally the driver of voluntary sustainability standards development was the intention to provide a tool that is independent from governments and can be used to achieve social and environmental change by using the markets in one country to affect production in another (Komives & Jackson, 2014). Marx and Wouters (2014) describe the uptake of voluntary sustainability standards 'at home' (in the food sector) as a reaction to new legislation attaching liability to the producers in order to guarantee (food) safety. In reality, the growth of voluntary sustainability standard systems can be observed by examining the quantity of standards over time. Figure 1 illustrates the number of voluntary sustainability standards since 1940, showing a steep increase around 1990, leading to around 400 different standards today.

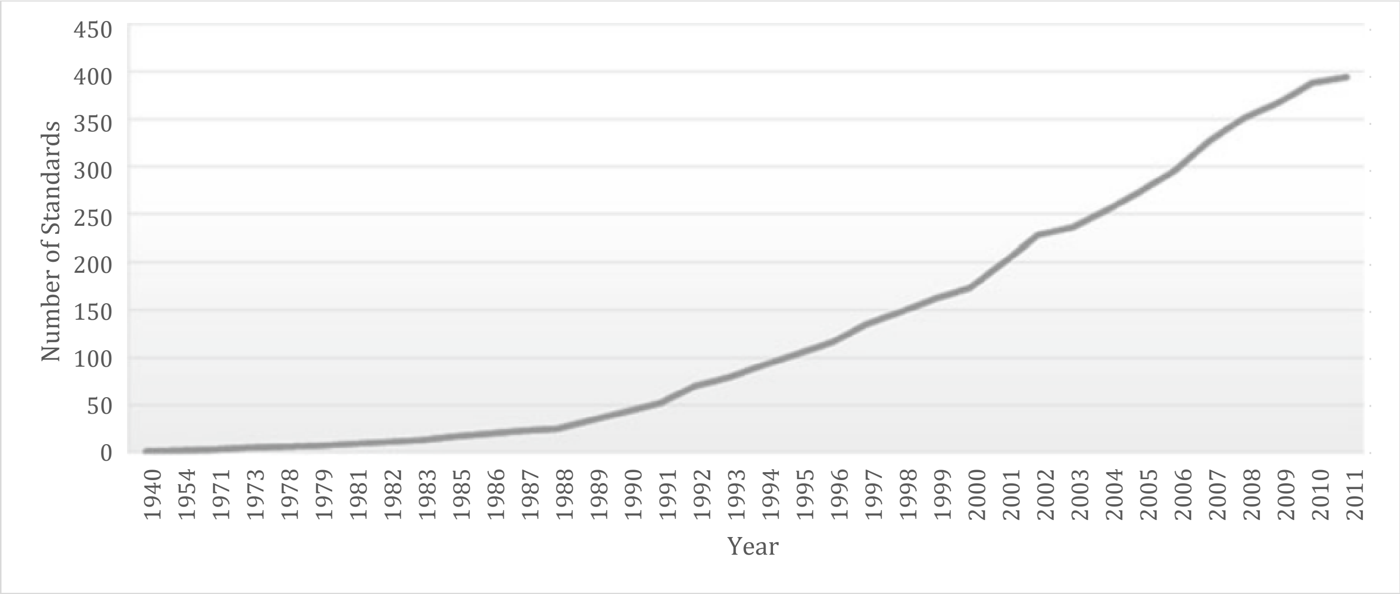

Figure 1 - Total number of voluntary sustainability standards by year of launch from 1940 to 2011 (Source: Komives & Jackson (2013) (Ecolabel Index 2013))

Complementary to this, Figure 2 shows that very different sectors now promote these standards and not just the more traditional ones where the environment is affected directly and where standards began to emerge in the past, such as forestry and agriculture. Standards can cover more than one sector, which explains the greater quantity shown in this figure.

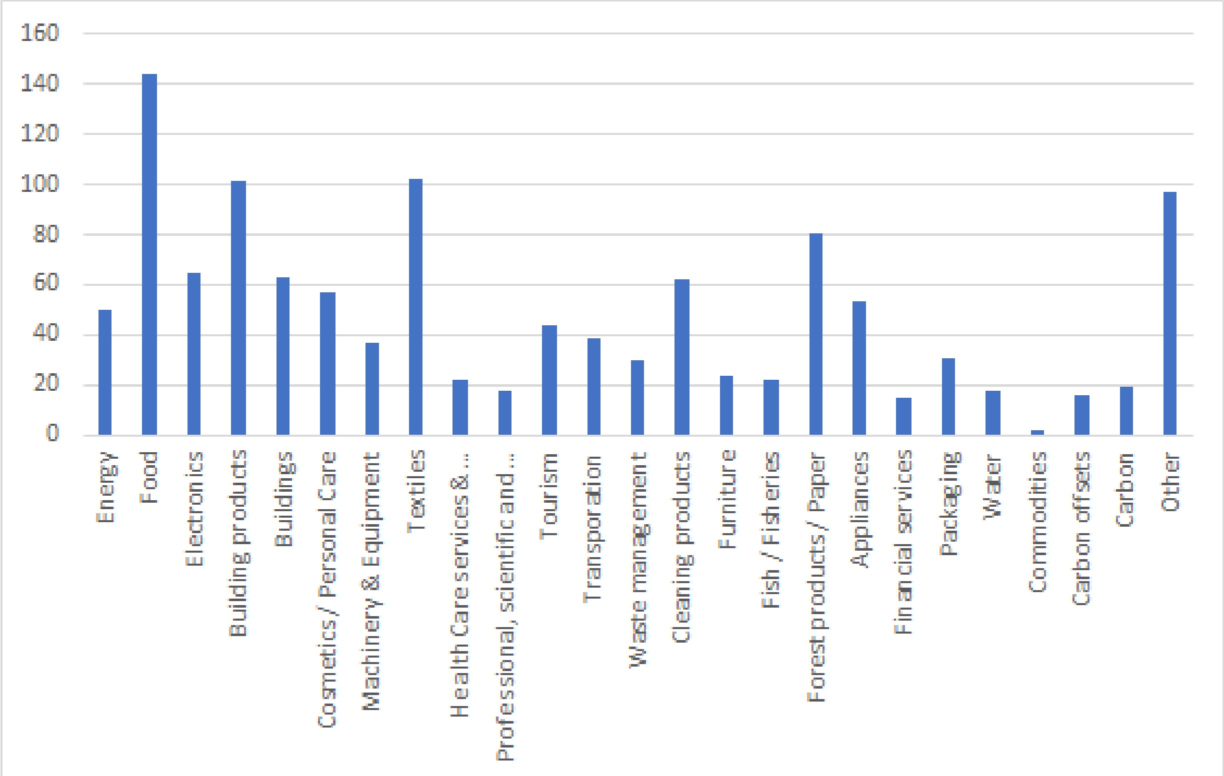

Figure 2 - Voluntary sustainability standard distribution across economic sectors (Source: Marx & Wouters (2014) (Ecolabel Index Database (N = 1218)))

However, when examining the historical development of voluntary sustainability standard systems, it is visible that not only the quantity increased, but also the quality of standards improved. The first private standards emerged in the first half of the 20th century in form of private organic standards on the local level with diverse criteria and practices, later rising to the national level to cover more sectors. The International Federation of Organic Agriculture Movements was established in 1972, presenting a shift from such local organic standards to an international one. Following the Max Havelaar standard in 1988, the fair-trade standards developed across Europe until Fairtrade International (called Fairtrade Labelling Organisation International then) was established in 1997. Around this time, in the 1990s, standards also began to seek a global consensus on sustainable practices by cooperating with one another. The formation of the ISEAL Alliance in 2002 is a result of such cooperation. Around 2000 the novel commodity-based roundtables emerged (e.g. Roundtable for Sustainable Palm Oil), where government, NGOs and industry can get together to develop standards for commodities with recognised environmental impacts (Komives & Jackson, 2014).

Nowadays, new trends are on the rise with standards using different models and assurance mechanisms. Traditionally standards require certain practices or management systems, but recently standards have begun to test for performance and results directly instead of relying on the assumption that those requirements automatically lead to the desired outcomes. Furthermore, certification that only judges producers with a pass or fail is no longer the only option as there are some schemes that award different degrees of certification (e.g. Green Building Council; platinum, gold, silver etc.). Aside of that, cooperation among standards organisations has improved with some standards entering partnerships in which they share resources and/or information in order to improve their impact.

Regarding the assurance process, auditors, nowadays, are not necessarily accredited through a traditional national accreditation body, but more often through larger international accreditation bodies. Beyond this, auditors now provide ‘verification’, which has developed as an alternative to full certification. Verified producers do not receive the ability to use a claim or certificate, but through insights of the auditors, provided in the verification procedure, they may be able to improve their practices and move towards certification at a later stage.

Developments in technology also offer more options for the traceability of products, advancing beyond the ‘identity preservation’ (full traceability of products), ‘segregation’ (separation of non-certified and certified, but no tracking to origin) or ‘mass balance’ (keeping track of the certified share of the product) mechanisms (Komives & Jackson, 2014). Online systems can be used to, for example, track individual products or to provide more accurate representations of valid certificates.

Facing this recent rise in diversity and complexity of voluntary sustainability standards, a debate has emerged with critics warning of the implication of competition among standards and a risk of a ‘race to the bottom’ that could lead to standards with lax certification criteria simply to capture a greater number of producers. Bitzer, Francken and Glasbergen (2008) argue that the multiplicity of partnerships in the coffee sector led to parallel, competing sustainability standards, addressing slightly different issues and effectively preventing each other from controlling a significant share of the market, ending up in a situation where the mass-producers continue to determine the important parameters of the market. However, Reinecke, Manning and Von Hagen (2012) argue that ‘healthy competition’ can drive innovation and thereby increase the effectiveness of rulesetters. Actors who operate in the same standards market, despite having different strategies and standards, agree on a “common ground [and] baseline criteria for ‘legitimate standards’” (Reinecke et al., 2012, p.808), counteracting the ‘race to the bottom’ argument. Furthermore, competition can encourage innovation as it maintains the discussion on what is considered sustainable and, beyond this, an increased number of participants can “contribute to institutionalising sustainability certification [...] as was the case in sustainability reporting” (Reinecke et al., 2012, p.808).

Independent of whether competition is good or bad for the development of voluntary sustainability standards, it is apparent that evaluating standards is useful to discover which attributes allow them to grow and compete in the first place. Beyond this, creating clarity on which practices and attributes are favourable, would improve the performance of standard setting organisations. Another major benefit is that producers and consumers can be aided in their standard choice. A concept suitable for such a comparative evaluation of standard systems would appear to be the ‘good governance’ concept.

1.1.2 Basics of Good Governance

Good governance is a central concept of this paper. A majority of the normative literature on this concept debates the concept itself, its development and implications and interpretations on how good governance should function in various contexts (Mkandawire, 2007; Weiss, 2000; Doornbos, 2001; Esty, 2005), while empirical literature is a bit scarcer than that and usually still provides statements on how it ought to function (Phillips, Hailwood & Brooks, 2016; Aguilera and Cuervo-Cazurra, 2004; Zattoni & Cuomo, 2008; Woods, 2000). The concept can be applied to very different organisations ranging from codes of good governance for companies (Aguilera & Cuervo-Cazurra, 2004; Zattoni & Cuomo, 2008) and guidance for governmental institutions (Roy & Tisdell, 1998) to supranational (Esty, 2005) and international organisations like the International Monetary Fund and the World Bank (Woods, 2000). A look at these contexts reveals that good governance has not only been used to improve individual organisations but also to affect entire discourses. For instance, Phillips, Hailwood and Brooks (2016) engaged in an application of the good governance of oil in Ghana, based on the claim that oil as a so called ‘resource curse’ demands additional attention if the country in question wants to use it for development.

The exact criteria for what is considered ‘good’ governance can be quite exhaustive and usually vary, depending on the issue or type of organisation that is the current focus of the concept. For the purpose of this study it was necessary to define or adjust these criteria for the voluntary sustainability standard context, building on other literature.

1.2 Problem Definition

Examining the dimensions of ‘good governance’ has been subject of various studies, typically bound to organisations from different contexts. The voluntary sustainability standards and their network of organisations are somewhere in between all of these and yet have only been subject to individual assessments without much regard for good governance, despite becoming increasingly autonomous, independent from governments and industry, and thus increasingly able to govern their sectors. High quality governance in voluntary sustainability standard systems is important because it has several benefits to producers, the functioning of the system and the relationship between these two.

It is crucial to finally combine these two branches of literature and produce an assessment to find gaps and flaws in the governance structure of standard setting organisations in order to improve their performance and the use of voluntary sustainability standards as a governance tool towards a more sustainable society.

1.3 Scientific Relevance and Connection to the Theoretical Debate

In the scientific community, individual standard systems have typically been assessed with a focus on the efficiency or effectiveness of the standards they promote. Examples are cost-benefit analyses with the purpose of advising governments or producers or quality assessments of the standard's indicators or studies that assess the societal, economic and environmental impacts (Blackman & Rivera, 2011; Thorlakson, Hainmueller & Lambin, 2018; Tayleur et al., 2018). The quality of the organisations themselves is rarely examined by outsiders. While there are case studies examining standard setting organisations (Thorlakson, Hainmueller & Lambin, 2018) and good governance in various contexts (Aguilera & Cuervo-Cazurra, 2004; Zattoni & Cuomo, 2008; Roy & Tisdell, 1998; Esty, 2005; Woods, 2000), few make an attempt to combine the two discourses (Castka & Corbett, 2016).

Castka and Corbett (2016) tried to discover a causal relationship between the spread of voluntary sustainability standards and media coverage, stringency and governance. While they did find that governance positively affects the spread of a standard, they did not assess governance within the standard system, as their judgement was based on an elaborate survey with 67 experts from different backgrounds. Therefore, Castka and Corbett (2016) also fail to provide an objective method for the assessment, which appears to be non-existent throughout literature. An emphasis is put on the attempt to create a more objective method, since the majority of good governance assessments, for example the Millennium Challenge Account, Global Governance Initiative, Latino Barometer and Polity IV, mainly rely on subjective methods (usually surveys) (Hood, Dixon & Beeston, 2008; Besançon, 2003; Knoll & Zloczysti, 2012).

The Institute for Multi-Stakeholder Initiative Integrity (MSI Integrity) has developed the MSI Evaluation Tool for multi stakeholder initiatives, which incorporates some relevant good governance dimensions but also a great number of aspects that extend beyond the reach of good governance. This tool likely is the best attempt at combining the two mentioned discourses.

However, by providing a method that is able to specifically assess good governance of voluntary sustainability standards, research can continue to create tailored solutions to enhance good governance for individual and multiple standards or potentially the entire market of voluntary sustainability standards. On top of this, refining the features in the standard systems that are relevant to good governance will bring about their intended societal benefits and allow for further development of these systems.

1.4 Societal Relevance

Voluntary sustainability standards inform actors and ease the process of making product choices from the total variety of offered products for producers and consumers and, although sometimes disputed, lead to more sustainable practices. At the same time, producers benefit from market access, higher profits and assistance in day-to-day activities. According to Denny (2017), reputation, which can be enhanced through certification, has become a strong driver and allows a firm to have an increased mark-up price for products and services as well as attract outstanding employees, thereby achieving a superior financial performance.

Not only do producers benefit from the establishment of voluntary sustainability standards, but also the governmental and intergovernmental institutions that attempt to govern the market. Compliance with standards can lead to a better and more effective management of the production and distribution, effectively making use of more sustainable methods on a local level and thereby contributing to larger goals (Denny, 2017).

These benefits to the market and governments, and thereby ultimately the citizens, are enhanced if the system is well governed. Good governance of the standards also means that they may enjoy a variety of benefits themselves. Well governed voluntary sustainability standards have a lower risk of accidentally certifying low-performing firms or failing to remove certification as well as ensuring that producers fulfil their obligations, thereby increasing the credibility of the voluntary sustainability standard (Castka & Corbett, 2016). Furthermore, this allows the creation of trust between the standards, producers as well as the consumers and leads to an improved network of partners, implying an easier maintenance of relationships and resources necessary to the survival of the system (Castka & Corbett, 2016). Aside of this, the functioning of the standard system improves and thereby amplifies all the usual benefits that voluntary sustainability standards are supposed to provide. Beyond this, being able to objectively compare and assess standards, identifying favourable attributes, allows producers and consumers to more easily pick standards that suit their own needs.

Therefore, the societal benefit of this research largely lies in the benefits that the voluntary sustainability standard systems as a whole already do/attempt to provide to society. With the successful development of an assessment framework the barriers to good governance in individual standard systems should become identifiable and can then be removed in the future, enabling a better governance and thereby amplifying these various benefits. Logically, these effects will only occur if the decision-makers are inclined to review this assessment as well as its method and take it seriously.

1.5 Research Objective(s)

The main objective of this paper is the development of a framework that can be used to objectively assess the ability of standard systems, from any sector, to undertake good governance. This framework is developed based on a literature review of voluntary sustainability standard systems and their development as well as good governance theory and case studies and/or assessment methods, where good governance is operationalised. The second objective is the actual application of the assessment framework to a limited number of standards from different sectors. Such an application should improve the framework and check if it is possible to create a single framework for different standards and still derive meaningful comparative conclusions. Underlying these two objectives is the need to first synthesise literature on the current background and development of the voluntary sustainability standard systems as well as good governance, discovering the applicable good governance criteria for the voluntary sustainability standard context. With a basis from these two branches of literature, the objectives can be pursued. In the end this assessment framework should be a first step towards bridging the

1.6 Research Questions

Making the creation of an assessment framework for good governance in voluntary sustainability standard systems the main objective leads to the following research question with the standard system as the central object of study:

How can good governance of voluntary sustainability standard systems be analysed?

Related sub-questions for the development of (theoretical) background knowledge are:
1. How are voluntary sustainability standard systems organised?
2. What is the theoretical reasoning behind good governance?
3. What are the key aspects of good governance?
4. What is good governance in the context of voluntary sustainability standards?

Beyond these the next two sub-questions form a basis for the development and testing of an assessment framework of good governance in voluntary sustainability standard systems:

5. How can the key aspects of good governance of standard setting organisations be translated into feasible data collection/assessment methods?
6. Can the required data collection/assessment methods be combined and valued in a framework to measure the degree of good governance of standard setting organisations?

2. Theoretical Background

The theory used as a basis for the development of the assessment framework originated from two branches. Firstly, good governance in general as well as its development was examined. Followed by that, applications of good governance in various contexts, alongside examples, were identified and are presented below. Secondly, literature on voluntary sustainability standards was used to identify how the relevant organisations function, which elements are crucial and to what extent they show similarities and differences to organisations and institutions presented in case studies and other operationalisations of good governance from the private and public sector. Once these elements were sorted out, criteria for good governance in standard systems could be identified through a review of various initiatives and then used for the development of the assessment framework.

2.1 Theoretical Foundation of Standard Systems

The underlying motivation for the use of standards being the sustainability effects does not accurately explain the development of this phenomenon. Ponte and Cheyns (2013) point out that attempts to adapt to new commercial opportunities, changing consumption patterns and a demand for varying levels of quality has made the governance of commodities, trade and consumption more prone to constant change. In an addition to this, Marx and Wouters (2014), describe that it has become difficult for governments, especially in the food sector, to keep track of the risks associated with the rapidly growing quantity of products in their domestic markets.

Therefore, guaranteeing a certain quality has become increasingly difficult, resulting in popularity of credence claims, especially for attributes that are not physically visible. The truthfulness of these claims is maintained by a “certification, auditing and accreditation industry” (Hatanaka et al, 2005; Loconto and Busch, 2010; Mutersbaugh 2005; as in Ponte & Cheyns, 2013, p. 2), which uses voluntary sustainability standards to lift certified products above others by indicating a higher quality through labels (Denny, 2017). In certain sectors, especially agriculture and food, certified products have gained a considerable market share (Ponte & Cheyns, 2013), making certification a requirement to access these parts of the market.

Beyond this increase in complexity, global value chains are becoming more and more dominant as they “tilt the balance of power away from the public sphere to a more market-driven network, empowering new international actors” (Denny, 2017, p. 2). While this change is not necessarily bad, it did create issues (e.g. climate change, transnational crimes) that demand cooperation and undermine the sovereignty of states by creating interdependence (Denny, 2017). In the wake of this change, new forms of cooperation with different approaches have gained a foothold and often address the failure of traditional legal system to deal with non-country-specific interests, multiple-factors and transnational issues (Denny, 2017).

Complementary to the historical development provided in the background information (chapter 1.1.1), Marx and Wouters (2014) identify three theoretical perspectives that suggest drivers of the emergence of voluntary sustainability standards. The first perspective suggests that voluntary sustainability standards are a tool of NGO-driven institutional capacity building, claiming that NGOs increasingly use voluntary sustainability standards as a co-operative strategy to affect the market instead of relying on confrontational methods. Another perspective determines that the driver is the consumer demand for ethical products and the attention NGOs bring to unethical aspects of production and consumption, which lead to the adoption of voluntary sustainability standards by firms who seek to avoid confrontation, please customers and access niche markets (Marx and Wouters, 2014). The last perspective promotes that the driver is the adoption of voluntary sustainability standards as mechanism to pre-empt the establishment of government regulation, meaning that private firms prefer the voluntary adjustment to a certain private voluntary standard to the chance of an introduction of mandatory public standards.

2.2 Development of Good Governance

The origin of the concept of good governance and reasons for its burst in popularity on the international level around 1990 are not perfectly clear. Mkandawire (2007) explains that it was inspired by African scholars and has grown as a result of the UN's statement that the development in Africa is a governance crisis. Similarly, Doornbos (2001) relates good governance as a tool for development aid to the fall of the Berlin Wall, pointing out that those who provided support had no intention of helping those who conducted, what they considered, bad governance, thereby creating new political conditions for aid. Weiss (2000) summarises that it simply developed as a concept applied wherever development aid was requested. Either way, good governance now is commonly found in various contexts, used as a tool to evaluate and/or improve the governance of countries, provinces, cities, institutions and also private organisations like NGOs or companies. The implication that governance can be 'good' is rather novel and states that it is possible to judge 'how' an entity is governed, regarding procedures, transparency, quality, decision-making and more (Doornbos, 2011).

Bøâs describes good governance as the natural opposite of bad governance, which has been operationalised by the World Bank as "personalisation of power, lack of human rights, endemic corruption and un-elected and unaccountable governments" (as in Weiss, 2000, p.801). Complementary to this, Mkandawire (2007) states that good governance is achieved if it sustains a triad made up of (1) development, meaning a balance of economic growth, structural change and usage of available resources in a sustainable and efficient manner, (2) citizen and democratic rights and (3) social inclusion. According to Weiss (2000), good governance in a Western-style democracy requires the following attributes; "universal protection of human rights; non-discriminatory laws; efficient, impartial and rapid judicial processes; transparent public agencies; accountability for decisions by public officials; devolution of resources and decision making to local levels from the capital; and meaningful participation by citizens in debating public policies and choices" (Weiss, 2000, p. 801).

Since the concept is quite broad, a conceptualisation that is tailored more specifically to the evaluation of the performance of standard setting organisations was needed. Based on empirical literature (see Introduction; Aguilera & Cuervo-Cazurra, 2004; Zattoni & Cuomo, 2008; Roy & Tisdell, 1998; Esty, 2005; Woods, 2000), this conceptualisation had to be a synthesis of the most suitable elements of good governance from the three different sectors of society, described by Edgar, Marshall and Bassett (2006); government, civil society and private sector (see Figure 3).

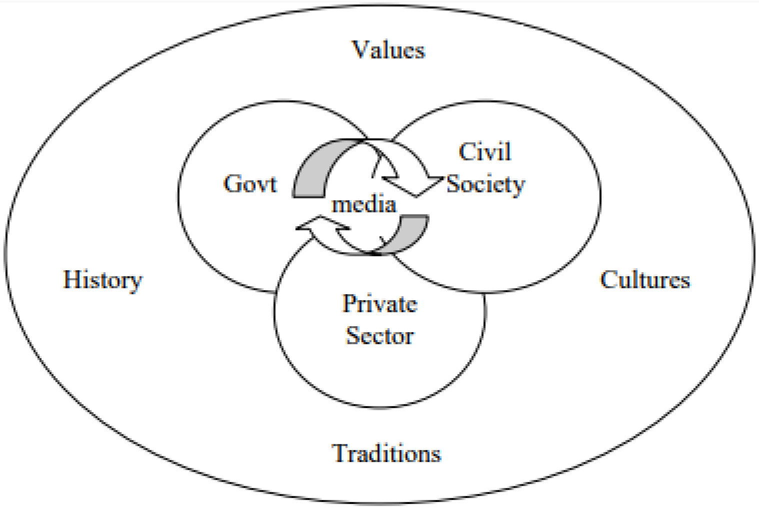

Figure 3 – Governance at the national level with actors occupying the social and economic landscape (by Edgar, Marshall and Bassett (2006))

Organisations cannot always be put into a single category as there may, for example, be government-funded NGOs or companies bound to the government by concession contracts. Standard setting organisations are expected to have strong ties to civil society but can belong to the other societal sectors based on their attributes. For instance, a standard could have its origin in the private sector, being established by a set of companies, is operated by NGOs and the scientific community and used by the government as a baseline for certain products. This mix makes it difficult to determine universal criteria for good governance in standard systems.

2.3 Theoretical Foundation of Good Governance

In the mid-1990s the UN contributed substantially towards shifting the debate around the model of good governance into a direction that includes the state rather than shoving it aside as has been quite common with the rise of neoliberalism (Weiss, 2000). The conceptual model provided by the UN provides a good first overview of the relevant principles (Figure 4).

The exact criteria of the good governance principles can be rather diverse, since it is difficult to measure the concept in reality and because they must be context specific in order to provide meaningful insights. In a policy brief, Graham, Amos and Plumptre (2003), expand and explain the underlying principles, pointing out that there is “strong evidence that these UNDP-based principles have a claim to universal recognition” (Graham, Amos, Plumptre, 2003, p. 3). It is important to note that these principles are related to the concept of good governance as a whole, meant to be applicable to any form of collective action. Graham, Amos and Plumptre (2003) explain that governance in general may occur for issues in the ‘global space’, ‘national space’, organization space’ or ‘community space’.

Figure 4 - Conceptual Model of Good Governance of voluntary sustainability standards (Source: http://www.unescap.org/sites/default/files/good-governance.pdf)

The global space includes large issues that go beyond the scope of individual nation states, while smaller issues, limited to the context of a single country, belong to the national space. Private and public organisations operate in their own context namely the organisational space. On the local level, where organisations do not possess a legal form without a formal governing board, the community space can be found (Graham, Amos, Plumptre, 2003).

Eventually, Graham, Amos and Plumptre (2003) split the UNDP principles into five good governance principles, valid for all actors in all spaces; 1. Legitimacy and voice, 2. Direction, 3. Performance, 4. Accountability and 5. Fairness. If these nine aspects (see Table 1) are fulfilled and thereby the five principles maintained, governance can be judged to be good. These principles are fundamental for the rest of the paper.

Box 1: Five Principles of Good Governance	
The Five Good Governance Principles	**The UNDP Principles and related UNDP text on which they are based**
1. Legitimacy and Voice	**Participation –** all men and women should have a voice in decision-making, either directly or through legitimate intermediate institutions that represent their intention. Such broad participation is built on freedom of association and speech, as well as capacities to participate constructively. **Consensus orientation –** good governance mediates differing interests to reach a broad consensus on what is in the best interest of the group and, where possible, on policies and procedures.
2. Direction	**Strategic vision –** leaders and the public have a broad and long-term perspective on good governance and human development, along with a sense of what is needed for such development. There is also an understanding of the historical, cultural and social complexities in which that perspective is grounded.
3. Performance	**Responsiveness –** institutions and processes try to serve all stakeholders. **Effectiveness and efficiency –** processes and institutions produce results that meet needs while making the best use of resources.
4. Accountability	**Accountability –** decision-makers in government, the private sector and civil society organizations are accountable to the public, as well as to institutional stakeholders. This accountability differs depending on the organizations and whether the decision is internal or external. **Transparency –** transparency is built on the free flow of information. Processes, institutions and information are directly accessible to those concerned with them, and enough information is provided to understand and monitor them.
5. Fairness	**Equity –** all men and women have opportunities to improve or maintain their well-being. **Rule of Law –** legal frameworks should be fair and enforced impartially, particularly the laws on human rights.

Table 1 - Five Principles of Good Governance by Graham, Amos and Plumptre (2003) based on the 1997 UNDP good governance principles.

3. Methodology

3.1 Research Strategy

The research can roughly be divided into five stages (see Figure 5), although there are overlaps and from time to time previous stages had to be reviewed. After (theoretical) background information (stage 1), initiatives and research that applies or analyses was collected, good governance could be examined (stage 2). Based on this information the assessment framework was created (stage 3) and then tested on selected cases (stage 4). In the end these assessments were evaluated and used to optimise and critically review the assessment framework (stage 5).

Information on the UNDP good governance principles was used as a starting point for the research. Standard systems were examined in a more general manner with neutral background information and then evaluated regarding how the different UNDP good governance principles function in the context of standard systems. During this evaluation, similarities and differences to the private and public sector were pointed out as it was fundamental for the next step of the framework. In the literature review a great number of existing initiatives and research was examined with a focus on how aspects of good governance or good governance as a whole is operationalised and analysed. Since the quantity of research that combines good governance with voluntary sustainability standards was rather scarce, experiences from the public and private sector were crucial to this task. Criteria and indicators, used by initiatives and researches that evaluate the private and public sector, were extracted and compiled for a potential adjustment and inclusion into the framework for standards. However, not only were these initiatives used to discover criteria, but they also provided a sense for how such an assessment framework could be made applicable and usable.

The research framework was to be based on all of the information found prior to its creation and had to go through several steps of refinement. A large number of measurement methods for every UNDP principle was available at this point, meaning that the major task was the downsizing of a very large set of criteria as well as the adjustment of these criteria so that they may be applicable to the standard system context. During this process it was important to constantly review the information on voluntary sustainability standards and also check the applicability of certain criteria if the existing information provided no certainty.

Finally, the assessment rounds presented as much of a test run for the framework as they did an actual assessment of the selected standard systems, as it was expected that the framework will still be rather incomplete and lacklustre. The first round of the assessment was based on data that could be found online, while the second round was supposed to be conducted through communication with experts from the standard organisations.

A division into two assessment rounds was made for a couple of reasons. By conducting a first round that was exclusively based on publicly available data the ‘transparency’ aspects could partially be judged, making a high availability of information a criterion of the framework. Besides this, being able to provide the experts with a rather complete sets of data ought to allow them to provide more detailed feedback with less of their time spent on criteria that may be rather easy to answer without their input. It was expected that the people working at the standard organisations have very limited spare time available. Another reason was that potential errors in the framework could be spotted during the first round, before having the experts review an aspect that might not make any sense to them.

At last, the assessments were reflected upon and, in combination with the first and second assessment rounds, used to critically review the framework and suggest improvements that should follow.

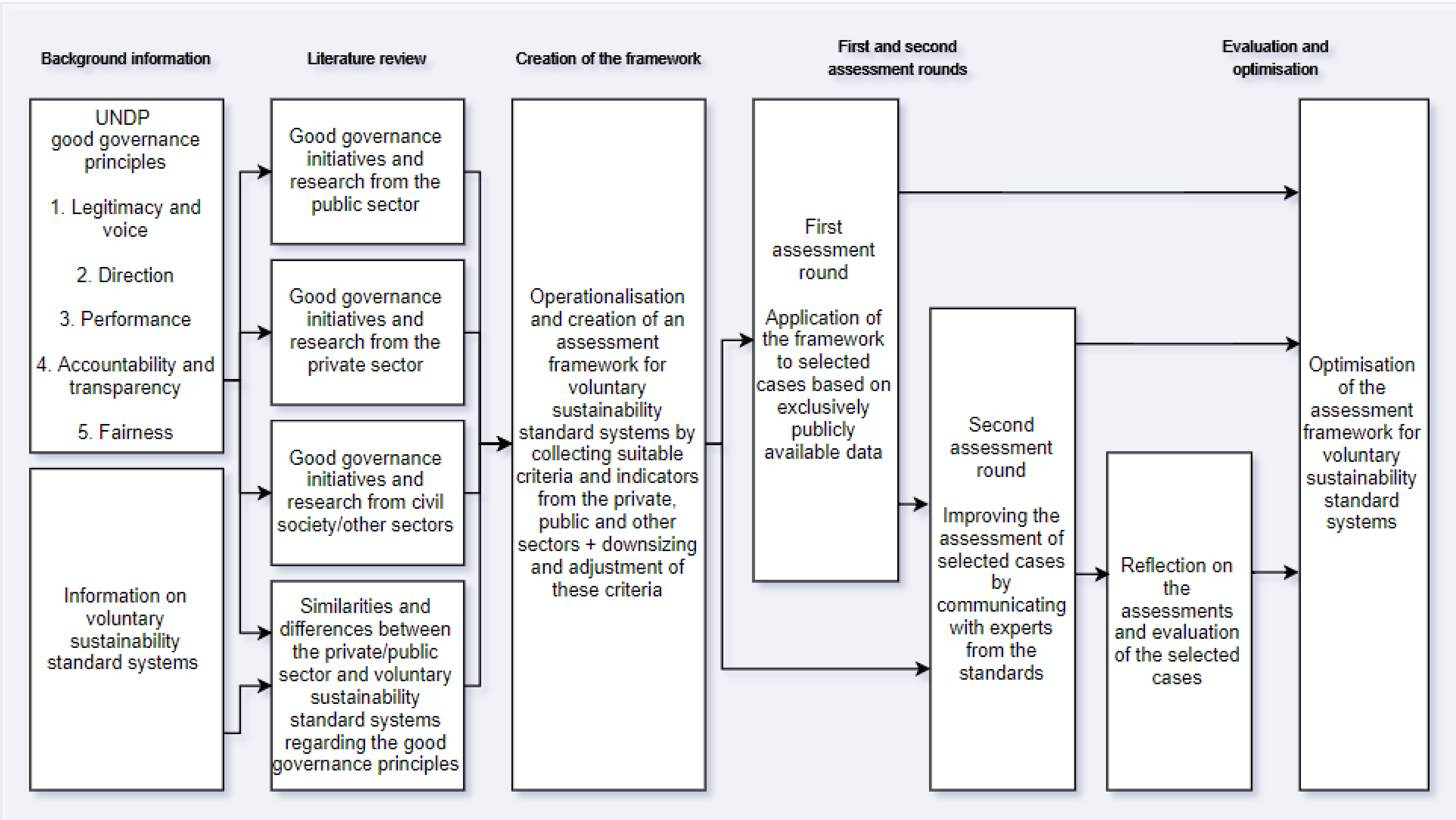

Figure 5 - Research framework of the study

3.2 Data Collection

In the first half (stage 1-3) of this research, data collection mainly consisted of the literature review, obtaining information from diverse journals found through online libraries/search systems such as Scopus, Wiley Online Library et cetera. These peer reviewed articles stemmed from research on good governance and voluntary sustainability standard systems. Another crucial source was information on existing initiatives, stemming from articles or simply their websites, that examine, consciously or not, aspects of good governance in the public, private or civil society/other sector. This included reports, templates, evaluation forms, methodologies et cetera that are used by the initiatives to conduct their own assessments. At this point, the websites of standards were occasionally visited to find additional information whenever the background information on the functioning of standards, based on the peer reviewed articles, did not provide sufficient insight to determine which measurement methods were appropriate for this research.

During the second half of the study (stage 4 and 5), an entirely different data collection process took place. As pointed out earlier, in the first assessment round, data collection was limited to publicly available data in order to check for transparency and prepare a dataset for the experts. Depending on the indicators and criteria developed for the assessment framework, quantitative and qualitative data was collected by examining the standard setting organisation websites, official reports, documents and statements of the selected voluntary sustainability standards. Any source that could be directly related to the standard system, clearly proving that information originates from those organisations was seen as usable.

Since it was expected that a lot of the information required to fill out the assessment framework would be non-disclosed or incomplete, the second assessment round was used to complete the assessment of the selected standards. Standard setting organisations were approached and requested to provide additional information to fill any gaps left by the first assessment round as well as confirm unclear information. This communication was planned to occur through interviews, where experts from the organisations could be consulted about individual criteria and the framework itself. These people were expected to have a very good understanding of the functioning of standard systems as well as the achievement of various good governance aspects in that context and thus should be able to provide great insights. After completing the first assessment round, experts from the standard systems were provided with a version of the assessment framework that included all information that was gathered up to this point. The difference to the version of the assessment used in the paper (see Appendix 1) was that the 'availability' and 'score' column were replaced by the 'user guide' and another column to ask specifically what information would be valuable for the criterion in question. In an addition to this, the experts, if able to provide additional information, were supposed to explicitly point out whether this information is publicly available and was thus just not found by the author during the first assessment round. This was important to evaluate the transparency aspect of the framework.

However, the second assessment round was not as successful as expected with only little actual feedback and no time to conduct the communication in the preferred manner. Mainly because of time constraints, the idea of conducting an interview was dropped in favour of requesting the experts to just fill in the provided assessment form.

3.3 Case Selection

Case selection focused on medium-sized or larger voluntary sustainability standards, because it was assumed that those have the capacity to provide information if requested during the second round of the assessment phase. Smaller organisations were not considered during this research, since data collection would be more likely to require a greater input of time as well as more communication with experts from those organisations. Focusing on 'voluntary' sustainability standard also implied that mandatory standards were not considered. The selected standards were supposed to operate in different sectors to ensure that the applicability of the framework would be more universal and thereby usable as a comparative tool.

The assessment covered the coffee and forestry sectors (see Figure 6), examining four organisations from these two different sectors. From each of these sectors one voluntary sustainability standard was picked that operates since the early days of standard setting in that particular sector, being a frontrunner (FSC and UTZ/RA), and one that was created as a follow-up (PEFC and GCP).

Figure 6 - Selected voluntary sustainability standards from the forestry and coffee sector

From the forestry sector the most popular and largest voluntary sustainability standards were selected, being the Forest Stewardship Council (FSC) and the Programme for the Endorsement of Forest Certification (PEFC). The FSC was founded in 1994 and with its slogan 'Forests For All Forever' seeks to "promote environmentally appropriate, socially beneficial, and economically viable management of the world's forests" (FSC, 2018). The standard itself demands attributes that create a more sustainable production and supply chain. However, the objectives of the standard go far beyond just creating a sustainable supply of forest-based product, since empowering indigenous people and maintaining ecosystem services also are part of their core values.

Although there still are great distinctions, the PEFC, created in 1999, tries to realise rather similar goals. The PEFC is "committed to conserving forests and their invaluable biodiversity, and the communities and, families that own, work and live in and around forests" (PEFC, 2018). While the PEFC certifies a larger quantity than the FSC, it may be argued that this is due to differences in the certification requirements and goals that these organisations try to achieve. A major difference between these two standards, beyond the goals and standard requirements, is that the PEFC is governed by its General Assembly which has roughly 120 members, while the FSC appears to be more democratic with more than 1000 members eligible to participate in the decision-making processes at their General Assembly.

In the coffee sector UTZ Certified, which was established in 2002 and recently merged with the Rainforest Alliance (mostly referred to as UTZ/RA in this paper), was selected. although the focus of this paper will still be on sources related to UTZ. The focus of this standard is on sustainable farming and not simply coffee. Next to coffee, it also actively certifies other products such a cocoa and tea. Through an implementation of sustainable agricultural practices, gender equality, increased income and quality of life improvements for farmers, UTZ/RA seeks to create a better and more sustainable world. Like the PEFC and FSC, UTZ/RA's certification allows for the usage of their label on products.

The Common Code for the Coffee Community (4C), established in 2006 and later (April 2016) evolved into the Global Coffee Platform (GCP). The GCP is the only standard in this research that does not promote a label alongside its standard. Different from the other standards, the GCP consciously submits that it offers a verification that is relatively easy to achieve and hence sees itself as a starting point for the uptake of more demanding certification schemes. However, it also has sustainability as a core objective next to the facilitation of communication among coffee-sector stakeholders.

4. Analysis

4.1. The Functioning of Standard Systems

While these standard systems are voluntary, they can become binding if they are part of a private contract. Often, certification is required if producers wish to enter a certain (part of a) market, despite the absence of a legal enforceability of standards. Depending on the standard, up to four different organisations can be involved directly in the activities of the labelling initiative. In the scientific literature, the terminology for these different organisations can vary and therefore a brief description of the entire system is provided at this point.

Voluntary sustainability standard systems, also known as labelling initiatives and often referred to as 'standards', 'voluntary standards', 'standard systems', are composed of five elements which, together, create an incentive for actors to implement more sustainable practices (Komives & Jackson, 2014). These elements are the standard (targets, criteria, indicators...) itself, standard setting organisations, assurance systems, capacity building of producers and traceability of claims (Komives & Jackson, 2014).

A voluntary sustainability standard system always has its origin in a standard setting organisation. These entities, as described by the ITC, are in charge of developing the standard initiative, code of conduct or audit protocol and can further be classified as private (e.g. NGO, industry association, company), public (e.g. governmental agency) or international (e.g. United Nations, OECD, ISO) (International Trade Centre, 2018). The distinction between the highest/international unit, which typically retains the decision-making power and manages day-to-day activities, mostly including standard adjustments, and subsidiary units is one that is quite important for the assessment phase. Depending on the size of the scheme, the standard setting organisations can have multiple legal entities as there may be subsidiary entities, which may still qualify as a 'standard setting organisation', that overtake activities in a certain region/country. However, while not all standard setting organisations promote a label, or claim to show off to the end consumers, the bare minimum of what they bargain with is the permission to make business-to-business claims.

Certification, the result of the assurance process by which a producer is eventually awarded with the ability to make such a claim or attach a label to their products, is, usually, undertaken by an independent third-party auditor (Komives & Jackson, 2014). These auditors, often referred to as certification body, get in touch with producers and conduct inspections, quality tests et cetera in order to determine whether they fit the criteria of the standard of their standard setting organisation.

Whenever there is cooperation among the standard setting organisation, the auditors and the producer, the latter often has the opportunity to engage in capacity building activities. Sometimes such capacity building services are offered by the standard setting organisation itself but can also be undertaken by a partner organisation (then referred to as capacity building organisation) and aim for compliance of the producer with the standard.

The last and most intangible element is the traceability of the chain of custody of products, which can be determined to a varying degree by different kinds of traceability systems usually provided by the standard setting organisation. Traceability is important to provide proof, supporting the claim that has been made (Komives & Jackson, 2014).

Accreditation, although not touched upon by Komives and Jackson (2014), could be a sixth important element as it appears to also be crucial for the understanding and functioning of voluntary sustainability standard systems. Auditors often receive accreditation for the promotion of a voluntary sustainability standards through an accreditation body which may be related to the standard setting organisation but can also be completely independent. Through this accreditation, auditors gain a confirmation on their ability to provide independent and competent certification. For instance, the International Accreditation Forum claims that accreditation based on their ISO/IEC 17021 standard ensures competence, consistency and impartiality of certification bodies (International Accreditation Forum, n.d.).

An example of the voluntary sustainability standard system of the Forest Stewardship Council, based on information from their official website (Forest Stewardship Council, 2018) is used to clarify the relationship between different organisations. The Forest Stewardship Council is a standard setting organisation which has an international unit and a great number of subsidiary national units and is a member of the ISEAL Alliance and provides various auditors with details on their voluntary sustainability standards used to certify wood. These auditors, among others including Siparila Oy, Brigl & Bergmeister GmbH and AENOR Internacional S.A.U., are accredited through the accreditation body called Accreditation Services International. Capacity building programmes of the Forest Stewardship Council occur on various levels but not through a separate organisation or on a regular basis but rather through independent projects facilitated by the standard setting organisation (FSC).

4.2 Good Governance in Standard Systems

Building on scientific literature, standard systems were compared to the public and private sector with respect to the five different UNDP good governance principles. By checking for similarities and differences, the relevance of certain kinds of assessment criteria from other contexts for the voluntary sustainability standards were determined, attaching a greater relevance to, for example, criteria that measure accountability in corporations if accountability in corporations is very similar to accountability in standard systems. This step was necessary, because the criteria and indicators of the assessment framework ought to be taken from existing, functioning assessments that use criteria that have proven useful, with priority given to objective criteria and a minimised number of criteria created without support from the literature review process (Chapter 4.3). The principles were examined separately after a broader reflection, based on the insights of Bernstein & Cashore (2007) and Dingwerth & Pattberg (2009).

Bernstein and Cashore (2007) describe so called 'non-state market driven governance systems', naming the FSC, MSC and others as examples, and discuss the political legitimacy of these systems. They determine five-characteristics of an ideal non-state market driven system which also fit the definition of voluntary sustainability standard systems. Firstly, the system does not derive its authority from the sovereign authority of a state, although states may be involved by providing financial support and affect their activities. While the state may also be responsible for the creation of such a system, it is not accountable to the state. Secondly, the involved actors in the system move towards collective goals and values in a dynamic manner, constantly working on adaptation, inclusion and learning. Bernstein and Cashore (2007) see a fundamental distinction towards traditional ecolabeling initiatives, which tend to be more static in their requirements. At the same time, non-state market driven systems are more democratic, open and transparent than "business-dominated public policy networks […] [and] most corporate self-regulation and many social responsibility initiatives" (Bernstein & Cashore, 2007, p. 3). Thirdly, authority of these systems comes from the demand of producers and consumers for compliance with the systems criteria. As fourth characteristic, the goal of the systems is to reconfigure markets in an attempt to address global problems that will otherwise not be addressed by the private sector. Lastly, standards are created and checked, non-compliance resulting in consequences for those who (voluntarily) participate (Bernstein & Cashore, 2007).

Dingwerth and Pattberg (2009) examine transnational rule-making organisations contributing to sustainability efforts, including the FSC, MSC, PEFC but also organisations that do not set a standard like the Coalition of Environmentally Responsible Economies, defining these organisations as transnational organisations that establish specific and (at least minimally) effective rules with goals towards social and/or environmental sustainability. This means that their findings are relevant for transnational voluntary sustainability standards, but not smaller ones. A primary output of organisations like the FSC are behavioural prescriptions, identifying rather precise criteria and processes with broader goals (Dingwerth & Pattberg, 2009). The World Commission on Dams provides the main frame of reference for large dams, the FSC principles and criteria present a benchmark for forest management and the Global Reporting Initiative created the leading standard for the accounting of non-financial performance of corporations (Dingwerth & Pattberg, 2009). Dingwerth and Pattberg (2009) point out that enforcement of their objectives/standards generally relies on “(quasi-) coercive pressures, rational incentives and appeals to social appropriateness”, effectively making use of what they call; “an instrumental ‘do X to get Y’ logic” (Dingwerth & Pattberg, 2009, p. 712), meaning that a certificate is awarded if their rules are being followed. Regarding the internal design of standard setting organisations, governing organs usually consist of “a board of directors that legally represents the organization and oversees its policies, a permanent secretariat that coordinates the organization’s day-to-day activities, and a larger stakeholder body that is most often assigned a consultative status within the organization” (Dingwerth & Pattberg, 2009, p. 713). Another element of transnational rule-making organisations is that their public communications focus on process rather than results, with the purpose of emphasising inclusiveness, transparency and accountability (Dingwerth & Pattberg, 2009). However, this is unlikely to be true for small standard setting organisations since a focus on process induces greater governance costs and less organisational autonomy (Dingwerth & Pattberg, 2009).

Based on the five characteristics described by Bernstein and Cashore (2007), the discussion of Dingwerth and Pattberg (2009) and additional literature the five UNDP good governance principles for standard setting organisations were aligned and compared to the public and private sector. The following chapters (4.2.1- 4.2.5) and the literature review (4.3) were created simultaneously. On the one hand, standard systems themselves present the civil society sector but one the other hand not much information was found during the literature review which was the reason, despite the original plan to include it, to discontinue a comparison to this third sector,

4.2.1 Legitimacy and voice in standard systems

The first UNDP principle includes the need for *participation*, meaning that everyone should have some direct or indirect influence on the decision-making, based on the freedom of association and speech and the ability to be constructive in their participation. *Consensus orientation* also is relevant as there usually are divergent interests, creating a necessity for integrative strategies and an adequate facilitation of interactions.

Legitimacy and accountability are taken very seriously by standard setting organisations and, along with effectiveness, are probably the most crucial to the functioning of the system. However, it is difficult to strictly align this principle with either public or private sector mechanisms as there are elements from both dimension. Authority of the voluntary sustainability standard system is not derived from the state, but, similar to companies, is based on demand from the public for their certification (Bernstein & Cashore, 2007). Voluntary standards intentionally cut themselves off from the state, which is the traditional source of legitimacy (Boström & Hallström, 2013), since they typically strive for goals that go beyond those of the state. In this context, standards present a mix between private companies who derive their legitimacy from the market and NGOs, who rely on moral authority (Boström & Hallström, 2013). Although the side to which they may lean can vary to a great extent, it is fundamental that the quality of a good product is maintained. This hints at a connection between the legitimacy, effectiveness and accountability (participation) UNDP principles. For standards this means that “they have to maintain that their organisations are appropriate arenas for standard setting, which include the appropriate people, structures, and routines for the standard development work. Moreover, they must maintain that the standards produced are novel, ethically important, useful, and needed in various ways – that there is a value-added component of the rules set by the organization" (Boström and Hallström, 2013, p. 100). Gulbrandsen (2008) emphasises this position in-between the private and public sector by pointing out that legitimacy in standard systems is more fluid, as there is an ongoing process that shapes and reshapes ideas and ideals about the modes of governance, dominated by transnational business consultants, NGOs and the media with the state and businesses being established components. Voluntary sustainability standards have to be more flexible and reactive to this process since they, unlike the private or public sector, maintain their rulemaking authority through an "answerability to salient constituencies in its organizational field" (Gulbrandsen, 2008, p. 568). However, precisely this answerability also creates similarities between legitimacy of the state with legitimacy of standards. While society should trust the state as legitimate ultimate authority for it to function, voluntary sustainability standards need stakeholders to support their system. Standards seek to gain such support through participatory mechanisms such as public comments for policy and stakeholder or expert consultation rounds (Dingwerth & Pattberg, 2009), creating a sense of ownership through cooperation that empowers the stakeholders (Boström and Hallström, 2013). Stakeholders who are enabled to voice their thoughts should be involved in a dialogue where responses can be given, and decisions can be explained if necessary. Since voluntary sustainability standards need/want to gain broad support, just like democratic states, transparency, public accountability and openness are deemed as important ideals.

Overall it appears that the basis for legitimacy, demand for the product, is very similar to that of the private sector, but its maintenance is mainly based on stakeholder support as is, partially, the way that states obtain legitimacy.

4.2.2 Direction in standard systems

Strategic vision is the core element of the UNDP 'Direction' principle, emphasising that actors must have a perspective and an idea of what is needed for long-term good governance and human development, based on historical, cultural and social complexities.

With 'change' being a core aim of standard systems, they distance themselves from most private organisations and align more with the goals of states. Most of the goals and targets that standards seek are of a collective nature, just like most goals in the public sector, yet, often the standards go beyond those, as they try to address issues where the public sector failed, and the private sector ignored its existence (Bernstein & Cashore, 2007). Beyond this, standard setting organisations tend to create precise criteria and methods to reach goals which usually are rather vague (Dingwerth & Pattberg, 2009). Although governments do occasionally create quantitative targets, precision in achieving those is uncommon and often not punished. Compare to this, companies usually do aim at more defined values, often expressed in terms of money and at the same time constantly reflecting on those aims during the process.

Governments, depending on the goal, and standard setting organisations work according to a so-called 'theory of change'. Brest (2010) defines a theory of change as the "empirical basis underlying any social intervention" and it is part of an organisation's "strategy, or logic model" used to achieve a certain impact (Brest, 2010, p. 49). Theories of change develop alongside their environment and should eventually demonstrate an impact through empirical validation (Brest, 2010). Testing for the soundness of a theory of change is useful if the impact is difficult to measure

(Brest, 2010) as is often the case with the goals that governments and voluntary sustainability standard systems try to achieve. Komives (2014), on behalf of the ISEAL Alliance, explains that a theory of change for standards consists of five building blocks completing one another; vision of success, pre-conditions (sub-goals), strategies or interventions, impact pathways and assumptions.

4.2.3 Performance in standard systems

Responsiveness simply means that the stakeholder interests must be addressed within a reasonable timeframe through monitoring, reporting, research and other activities. Aside of that, making an efficient use of available resources to reach certain goals is obvious in its import for the performance *(effectiveness and efficiency)* aspects.
Regarding the performance in comparison to the public and private sector a number of similarities can be found, despite the uniqueness of this aspect in standard systems. For voluntary sustainability standard systems, the effectiveness itself is not comparable to that of companies, where it usually is expressed in financial terms. Standard setting essentially is just one specific governance tool, which can actually be found in the public sector as well, making the work of standards quite similar to the public sector regarding the effectiveness aspect. Standard systems and governments seek to create (at least minimally) effective rules and behavioural prescriptions and can be evaluated based on how effective these rules are (Dingwerth and Pattberg, 2009). Yet, the tool kit of the public sector is far broader, without having to rely on a single governance tool, and beyond that, is supported by a far more solid base of authority and power with the ability to actually push actors more forcibly into the desired direction.
Responsiveness also appears to be taken more seriously by standards than the public or private sector. While there is lobbying and consultation in those contexts, standard setting organisations usually try to actively involve stakeholders in their system, although the extent of this involvement and the usage of their input vary (Dingwerth & Pattberg, 2009).

4.2.4 Accountability in standard systems

Accountability consists of *accountability* of the decision-makers in government, private sectors and civil society to the public and *transparency*, being the free flow of information. *Accountability* can be maintained through clear mandates and division of tasks but can differ depending on the type of organisation and decision. *Transparent* governance requires that enough information must be available about the processes themselves, but also to understand these.
This UNDP principle presents, like the legitimacy principle, a middle ground between the ways accountability in the public and private sector work. Voluntary sustainability standards are not accountable to the people, like the government, but at the same time are far more transparent, democratic and open than companies (Bernstein & Cashore, 2007). Officially, the standard is, just like a company, accountable only to its own organs (e.g. board of directors) (Dingwerth & Pattberg, 2009), yet, the presence of numerous stakeholders indirectly creates an accountability towards the outside world. Despite an accountability deficit towards the public, due to a lack of

public delegation and a tendency to resist the state and democratic control, standard setting organisations still have to respond to the pressures of those that they claim to represent (Gulbrandsen, 2008). Gulbrandsen (2008) emphasises that such outward accountability still strongly depends on how seriously the pressures from the stakeholders are taken, which, in turn, directly depends on the power that they hold. This means that stakeholders with little influence create less pressure on the standard setting organisation. Overall, voluntary sustainability standards accountability resembles what can be found in private organisations, but with an increased importance of stakeholders.

4.2.5 Fairness in standard systems

The last UNDP principle, fairness, seeks *equity* in form of opportunities to all men and women to improve or maintain their well-being. Beyond this, the *rule of law* aspect determines that everything must be based on a fair legal framework or a set of principles with impartial enforcement to increase fairness.

Regarding the equity part of this principle, it is rather difficult to connect to the perspective of the public sector, which has to address the people, since standards are not accountable to those. Other than that, it is reasonable to assume that employees in the organisation should have equal opportunities, rights and chances. Acting according to the rule of law must be compared to companies, as the voluntary sustainability standards has no authority to formulate legislation and therefore is only subject to existing law and perhaps an effort to prevent wrongdoing. Standard setting organisations do check for non-compliance and punish, to a limited extent, those participants that do not conform to what they agreed to (Bernstein & Cashore, 2007). Aside of that, the prevention of corruption is a reoccurring theme in scientific literature on the rule of law in companies.

4.3 Literature Review

At this point, based on the five good governance principles, initiatives, measures, assessment tools and criteria suggested by the scientific literature, are presented. All of which focus on the public and private sector or, if applicable, civil society or other sectors. The reflection on those initiatives and their precise criteria and indicators, in combination with the findings from the previous two chapters allowed the selection and creation of usable criteria and indicators for the assessment framework. In the following sub-chapters every principle was reviewed once again with an emphasis on assessments from the private sector, followed by assessments of the public sector. Initiatives are only presented if there was a potential, based on Chapter 4.2, that they could be applicable to voluntary sustainability standards.

Table 2 presents those initiatives and researches, selected because they explicitly or implicitly try to address good governance as a whole or aspects of it.

Public sector	Private sector	Other
1. Legitimacy and voice (participation and consensus orientation)		
Afro Barometer	Research by Wyk and Chege (2004)	Research by Bernstein and Cashore (2007)
Research by Arat (1991) – Democraticness	IoD Good Governance Index	Research by Dingwerth and Pattberg (2009)
The Database of Political Institutions	UN Global Compact Self Assessment Tool	Research by Saner and Wilson (2003)
Participatory Development and Good Government Ranking	Robecosam Survey	Research by Edgar, Marshall and Bassett (2006)
Research by Henisz (2000) - Political Constraint Dataset (POLCON)	EO100 Standard	
Polity IV		
Research by Vanhanen (2000) – Polyarchy		
2. Direction (strategic vision)		
Good Governance Guide (KPI)	Article by Wyk and Chege (2004)	Research by Saner and Wilson (2003)
		Research by Edgar, Marshall and Bassett (2006)
3. Performance (responsiveness and effectiveness and efficiency)		
Worldwide Governance Indicators	Research by Wyk and Chege (2004)	Research by Dingwerth and Pattberg (2009)
World Economic Forum's Global Competitiveness Report	IoD Good Governance Index	Research by Saner and Wilson (2003)
Good Governance Guide (KPI)	GoodCorporation Framework on Human Rights	Research by Edgar, Marshall and Bassett (2006)
Research by Feng, Kugler and Zak (2000)	Robecosam Survey	
Research by Bovaird and Löffler (2003)	Indicadores Ethos para Negócios Sustentáveis e Responsáveis	
Millennium Challenge Account	EO100 Standard	
4. Accountability (accountability and transparency)		
Worldwide Governance Indicators	Research by Wyk and Chege (2004)	Research by Dingwerth and Pattberg (2009)
Good Governance Guide (KPI)	IoD Good Governance Index	Research by Saner and Wilson (2003)
Country Assessment in Accountability and Transparency	EO100 Standard	Research by Edgar, Marshall and Bassett (2006)
Millennium Challenge Account		
Afro Barometer		
Bribe Payers Index		
Community Information, Empowerment and Transparency International		
Polity IV		
5. Fairness (equity and rule of law)		
Worldwide Governance Indicators	Research by Wyk and Chege (2004)	Research by Saner and Wilson (2003)
Economic Freedom of the World Index	UN Global Compact Self Assessment Tool	Research by Edgar, Marshall and Bassett (2006)
Freedom House	Robecosam Survey	
Bribe Payers Index	EO100 Standard	
Community Information, Empowerment and Transparency International		
Corruption Perceptions Index		
Research by Clague, Keefer, Knack and Olson (1999)		
Millennium Challenge Account		

Table 2 – Initiatives and research from the public, private and other sectors

The initiatives were split based on their potential to provide criteria for the assessment of the five UNDP good governance principles in the standards context. They were examined with regard to the categories they propose, while the exact criteria and indicators were reviewed and adjusted later, during the creation of the assessment framework for standards. Despite a dominance of initiatives using subjective criteria and methods, most of the selected initiatives and researches use objective criteria to draw conclusions on the performance of the public or private organisations. The entire set of initiatives originated from an elaborate search for assessments of good governance through web searches and scientific database searches (e.g. Scopus) that led to websites of existing initiatives that provided an insight into their methodology or scientific articles and initiative documents. Key words such as ‘good governance in companies/private or public sector/ government/ (international) institutions/ state/ country’, ‘assessment/ scoring system/ initiative/ evaluation’ and the principle names or synonyms thereof were used during this search. In an addition to this, sources that cited the UNDP principles and closely related documents were examined. Crucial for the discovery of these initiatives were a few articles that examined a great number of initiatives at once.

Besançon (2003) examined forty-seven data sets that mostly focus on public sector initiatives, while De Felice (2015) looked at thirty initiatives focusing on business and human rights. Another comparison was undertaken by Van de Walle (2006), who compared bureaucracies based on World Bank, European Central Bank, Worldwide Economic Forum and International Institute for Management Development assessments. Overall fourteen articles and additional unscientific source were examined that presented at least one initiative examining good governance in the public, private or other, mostly civil society, sectors (Arat, 1991; Bernstein and Cashore, 2007; Besançon, 2003; Bovaird and Löffler, 2003; Clague, Keefer, Knack and Olson, 1999; De Felice, 2015; Dingwerth and Pattberg, 2009; Edgar, Marshall and Bassett, 2006; Feng, Kugler and Zak, 2000; Henisz, 2000; Saner and Wilson, 2003; Van de Walle, 2006; Vanhanen, 2000; Wyk and Chege, 2004). Originating from these articles and other sources, twenty-seven initiatives were examined regarding the criteria they use. The initiatives marked green (thirteen) in Table 2 are the ones described in more detail below. However, a great number of these twenty-seven examined initiatives were eventually discarded and are marked red in the table. Some of these simply did not contribute any additional criteria. For instance, the Briber Payers Index criteria overlap a lot with those of the Worldwide Governance Indicator’s corruption aspects. Others, like the self-assessment tool the Country Assessment in Accountability and Transparency or the service of Community Information, Empowerment and Transparency International do not provide sufficient further information. Yet again, some other initiatives were too focused on a certain niche of their sector, like the Economic Freedom Index which focuses, among other aspects, on the judicial system, laws or specific issues like terrorism and crimes. The ‘other’ sources presented in the table present alternative approaches dealing with partnerships, the concept of stewardship or the civil society sector. However, none of these provide criteria for the evaluation of good governance, but merely discuss the concepts or broad criteria on a more theoretical level.

Public sector initiatives aligned rather well with their ideas of what good governance was supposed to be. Besançon's (2003) research provides an overview of good governance initiatives for the public sector and was used to discover most of the legitimacy, accountability and fairness initiatives. Good governance is described as the result of a nation-state providing a high quality of certain political (public) goods to those living within designated parameters, in an effective manner (Besançon, 2003). These public political goods are; "rule of law; political and civil freedoms; medical and health care; schools and educational instruction; roads, railways, the arteries of commerce; communications networks; a money and banking system; a fiscal and institutional context within which citizens can prosper; support for civil society; and a method of regulating the sharing of the environmental commons" (Besançon, 2003, p. 1). This definition showed a lack of an emphasis on the direction principle and the responsiveness part of the performance principle.

Knoll and Zloczysti (2012) review the indicators of the Millennium Challenge Account, pointing out that it does not provide any evidence for a direct relation of its governance indicators to the UNDP good governance concept. The Millennium Challenge Account uses seventeen initiatives, which are grouped into three policy dimensions; 'ruling justly', 'investing in people' and 'economic freedom'. Although the 'economic freedom' policy dimension does appear to be something of an addition, the Millennium Challenge Account's implicit understanding of good governance aligns very well with the performance, accountability and fairness principles, while scratching parts of the legitimacy and neglecting the direction one. It should be noted, that the Millennium Challenge Account, in an attempt to cover good governance more broadly, makes use of other existing and more precise initiatives, such as the World Bank Governance Indicators.

The Governance Institute of Australia (n.d.) takes a slightly different approach by bonding good governance with key results areas, making the achievement of these the basis for good governance within the institution. This approach heavily depends on the agreements, service charter etc. that control the institution, but also whether the key performance indicators are accurate. Beyond this, institutions conduct good governance if they review and improve their own key performance indicators. Aside of measuring the performance, key performance indicators ought to constantly channel the progress towards certain goals, thereby contributing to the direction principle. However, the legitimacy and fairness principle are not part of this view, while accountability is limited to the stakeholders' gaining a (more) quantitative tool for the review of the institution, if it is made publicly available.

As previously mentioned, the good governance definitions from the public-sector initiatives aligned rather well with the good governance concept of the UNDP, however, applications of good governance in the private sector often emphasised corporate governance and the creation of codes of good governance instead of the actual principles of the UNDP concept.

Codes of good governance are a set of best practices for the board of directors and other governance mechanisms embedded in the legislation of a country, commonly issued by the stock market or managers' associations (Zattoni & Cuomo., 2008) and therefore generally not suitable, despite favourable core aspects. Core aspects of these codes are "fairness to all shareholders, clear accountability by directors and managers, transparency in financial and non-financial reporting, the composition and structure of boards, the responsibility of stakeholders' interests, and for complying with the law (Gregory & Simmelkjaer, 2002; Coombes & Chiu-Yin Wong, 2004 as in Zattoni & Cuomo, 2008).

Van Wyk and Chege (2004) use and adjust the UNDP principles for the construction sector and thereby provide the only research (among the reviewed initiatives) that tries to broadly apply good governance as defined by the UNDP. The Good Governance Report's (Institute of Directors, 2017) criteria partly address the legitimacy, performance and accountability principles, but overall is more influenced by compliance with laws and regulations rather than any ideals originating from a specific good governance definition. Similarly, the Equitable Origin 100 Standard (Equitable Origin, 2012) addresses all but the direction principle, but avoids defining good governance, while their task is to evaluate (energy) projects with respect to governance and their protection of the environment and the people. De Felice (2015), the last source used to find private sector initiatives, examined factors related to human rights, which arguably do connect to many of the UNDP principles, yet do not paint the picture in its entirety.

'Other' literature did address the UNDP good governance concept more directly but did not provide details and remained rather vague. Saner and Wilson (2003) lean on the UNDP definition, elaborating on the stewardship concept as a governance process. Effective stewardship criteria are presented and compared to the UNDP principles, with a lack of a parallel for the strategic vision and equity aspect, but other additions that find no parallel among the UNDP principles. Edgar, Marshall and Bassett (2006) essentially do the same, but create parallels for the evaluation of partnerships

Altogether the great variety of views on good governance in the public and private sector provided a rather spotty coverage of the UNDP good governance principles, which was the reason why multiple initiatives had to be used to find suitable criteria. Table 2 clearly shows that none of the initiatives covered more than four of the five principles. In the following two sub-chapters the selected initiatives are presented with more detail in full-text as well as information boxes. The content of these boxes, which can be found in Appendix 3, may vary and is only supposed to provide details on the structure and/or categories of criteria and indicators used by the initiative.

4.3.1 Public sector good governance initiatives

Beck, Clarke, Groff, Keefer and Walsh (2001) describe the Database for Political Institutions (see Box1, Appendix 3), a tool used for the comparison of political parties among different countries, that has been applied to 177 different countries. The set of 108 variables provide details on "elections, electoral rules, type of political system, party composition of the opposition and government coalitions, and the extent of military influence on government" (Beck et. al., 2001, p.166) and makes use of objective and transparent criteria and indicators. This database is still being updated with its last update being in January 2016 for the year 2015 (Cruz, Keefer and Scartascini, 2016). The Database for Political Institutions 2015 Codebook (Cruz, Keefer and Scartascini, 2016) provides the updated list of variables used by this initiative and demonstrates that most variables, such as those about the military or political parties, are of little use in the context of standards and could therefore not simply be transferred to this study. However, the dataset takes seats and the balance of power of the parties as important factors, which are aspects that were translated to this context. Another interesting detail is that, probably due to its importance for transparency, a lack of information often results in a score of 0.

The Political Constraint Dataset (see Box 2, Appendix 3) was created by W. Henisz (2000) and measures institutional commitment with a focus on competitiveness and participation. Henisz (2000), based on country risk indices, creates this measure of political constraints with an emphasis on two elements: “the number of independent veto points over policy outcomes and the distribution of preferences of the actors that inhabit them” (Henisz, 2000, p. 5). Beyond this, Henisz (2000) provides a set of criteria to estimate growth which is then weighted against the other two factors to reach conclusions. Therefore, similar to Beck et. al. (2001), this research did not provide any criteria useful directly to the assessment of voluntary sustainability standards, but it showed the importance of veto power in the participation process, which holds relevance for stakeholder participation.

Marshall and Jaggers (2002) worked on the Polity IV Project (see Box 3, Appendix 3) and created a data set for the coding of democracy and authority characteristics of states and still is in use today. Just like the two data sets before this one, it works with a point scale and rather objective criteria. The data set makes use of various indicators and scores, being the ‘institutionalized democracy’ (4 indicators) and ‘institutionalized autocracy’ (5 indicators), which lead to the ‘combined polity score’, ‘revised combined polity score’, ‘regime durability’ and ‘polity persistence’. Since there is a strong focus on the evaluation of executive authority and the attainment of such, this assessment also is rather specific, since other contexts do not have to deal with executive authority the way states do. However, the component variables were transferable after adjustments. Marshall, Jaggers and Gurr (2017) use scales (3, 4 or 5 points, depending on the indicator) with a precise description for the evaluation of the ‘competitiveness of executive recruitment’, ‘openness of executive recruitment’, ‘constraints on chief executive’, ‘regulation of participation and competitiveness of participation’. For example, ‘regulation of participation’ ranges from ‘unregulated’ (1), over ‘multiple identity’ (2), ‘sectarian’ (3) and ‘restricted’ (4) to ‘regulated’ (5) (Marshall, Jaggers and Gurr, 2017).

The research of Vanhanen (2000) (see Box 4, Appendix 3) combined competition and participation to create a measurement for democracy, split into “degree of electoral competition, degree of electoral participation and a combined index of democratization” (Vanhanen, 2000, p. 251). Competition is measured through the share of total votes among the smaller political parties in parliamentary or presidential elections and participation depends on the amount of people who voted compared to the total population. The participation indicator, although rather simple, could be applied in the voluntary sustainability standards context by just recording the presence of stakeholders or other important actors during certain meetings.

The Worldwide Governance Indicators (see Box 5, Appendix 3), the indicators from 2009 produced by Kaufmann, Kraay and Mastruzzi (2009), report on governance in more than 200 countries, focusing on the following six dimensions; voice and accountability, political stability and absence of violence/terrorism, government effectiveness, regulatory quality, rule of law and control of corruption. The data is updated annually, and scores are based on dozens of sources with hundreds of indicators, evaluated with statistical methods (Kaufmann, Kraay, Mastruzzi, 2011). This initiative operates quite similar to the Millennium Challenge Account, because it makes use of existing assessments and data sources, without collecting data on its own. A difference is that the Worldwide Governance Indicators provide more detail on what concepts were assessed. For instance, the ‘Voice and Accountability’ category draws upon data from 21 of the 33 different data sources overall using 59 concepts (Kaufmann, Kraay, Mastruzzi, 2009). The categories and according indicators aligned very well with the fairness and accountability principles but actually provided useful information for every principle of the UNDP. Box 5 provides an example of the voice and accountability category of the Worldwide Governance Indicators. The Freedom House, which was initially listed in Table 2, is not described in more detail in this chapter as it is one of these, overall 33, data sources that provide data and indicators to the Worldwide Governance Indicators.

The World Economic Forum’s Competitiveness Report (see Box 6, Appendix 3) (Schwab, 2017) mainly analyses the efficiency of states and their economy through its Global Competitiveness Index, building on twelve pillars that emphasise features of the state such as the infrastructure (pillar 2), health and primary education (pillar 4) and market size (pillar 10). Each of these pillars is further divided into weighted categories that have their own indicators. Box 6 visualises the categories of the first pillar, which seems to be the most promising one for finding relevant criteria for the standard setting organisation context. It must be noted that this initiative relies on expert surveys, hence, if any indicator is transferred information on it will have to be gained in a different way.

Based on various initiatives, the Millennium Challenge Account (see Box 7, Appendix 3) merely presents the framework for the evaluation of various good governance aspects and is mainly interested in foreign policies and institutions, making conclusions for countries as a whole. It evaluates the three categories; ‘ruling justly’, ‘investing in people’ and ‘establishing economic freedom’, which forms the basis for an extension of (U.S.) aid. Knoll and Zloczysti (2012) point out that six of the seven indicators focusing on governance are presented in the ‘ruling justly’ category, hence this initiative was mainly important for the fairness principle.

Corruption in the public sector of different countries is evaluated in a rather subjective manner by the Corruption Perceptions Index (see Box 8, Appendix 3) on an annual basis. The final score between 1 and 100 is awarded based on an evaluation of experts from different sources (13 in 2017) who examine factors such as bribery, diversion of public funds, use of public office for private gain, legal protection for whistle-blowers, journalists and investigators etc. (Transparency International, n.d.)

4.3.2 Private sector good governance initiatives

Among the initiatives in the private sector, the objective measurements were less specific with a tendency to use nominal indicators. The 2017 Good Governance Report (See Box 9, Appendix 3) of the Institute of Directors (2017) is the third of its kind and evaluates 100 of the UK’s largest listed companies’ corporate governance performance. In doing so, it provided criteria suitable for the legitimacy, performance and accountability principles, most of which originate from its ‘board effectiveness’, ‘audit and risk/external accountability’ and ‘stakeholder relations’ categories. All of their criteria are answered and scored based on ‘yes/no’-answers, percentages or flat numbers and together form the score of a company. Although part of the report relies on a survey of stakeholder perceptions, useful objective criteria were numerous.

The UN Global Compact Self-Assessment Tool and the RobecoSAM Survey are only two out of about thirty initiatives examined by De Felice (2015), who provides an overview of the human and business rights initiatives, identifying opportunities and challenges in the process. Based on the descriptions of De Felice, these two initiatives were selected and examined more closely.

UN Global Compact is a corporate sustainability initiative with the mission achieve broader societal goals (e.g. Sustainable Development Goals) and to have businesses and stakeholders operate responsibly by making use of ten principles (United Nations Global Compact, n.d. a). The UN Global Compact Self-Assessment Tool (United Nations Global Compact, n.d. b) (see Box 10, Appendix 3) is an extension of this initiative and allows businesses to check themselves on whether they submit to the goals and targets of the UN Global Compact initiatives. It sticks to the five UN Global Compact principles with the addition of a ‘management’ category, mostly using nominal indicators (yes/no-questions) with the opportunity to add comments. The principles 1, 2, 3 and 10 held the most valuable aspects transferable to the voluntary sustainability standards context.

RobecoSAM’s (n.d.) Corporate Sustainability Assessment (see Box 11, Appendix 3) has very detailed and industry specific indicators and relies on the participants (businesses) to fill out an elaborate survey. When closely investigating the indicators used in the survey, it seemed like no extensive part of the criteria and indicators could be transferred. However, the social dimension, customer relationship management, economic dimension and anti-crime policy held some use for the legitimacy, performance and fairness principles, although the majority of indicators did not appear to address any UNDP good governance principle.

The Equitable Origin’s EO100 Standard (Equitable Origin, 2012) (see Box12, Appendix 3) is a standard of its own and provided a great number of useful criteria. The facilitator of the standard evaluates projects in the energy and other sectors, working for the protection of people and the environment, coupled with a strong focus on stakeholder involvement and good governance. The standard consists of six principles, each of those having six to sixteen different provisions with their own set of quantitative and/or qualitative indicators. The fulfilment of these eventually leads to certification. For the purpose of evaluating standard setting organisations different provisions were examined and a great number turned out to be have a potential to contribute to the measurement of the UNDP good governance principles. While the ‘indigenous people’s rights (principle 3)’, ‘climate change, biodiversity & environment (principle 4)’ and ‘project life cycle management (principle 6)’ principles were of no use to the purpose of this study, the ‘corporate governance, accountability & ethics (principle 1)’, ‘human rights, social impact & community development (principle 2)’ and ‘fair labour & working conditions (principle 3)’ principles held some valuable provisions and indicators for the UNDP legitimacy, effectiveness (responsiveness part only), accountability and fairness principles (see Box 12).

4.4 Combining Good Governance Initiatives with Standards

Reviewing the initiatives examined in the previous chapter (Chapter 4.3), it must be concluded that, despite providing great aid in the creation of most of the criteria, no single initiative, on its own, managed to provide a significant number of criteria and indicators that were directly transferable without any modification. The main reason for this is that the criteria of the initiatives are either superficial or too specific to assess the unique parts of the voluntary sustainability standard systems such as, for example, the interrelationship between the different organisations in the system. Besides this, the definitions of good governance vary greatly, especially in the private sector, causing many criteria to measure aspects that might be irrelevant to the UNDP good governance definition that is used in this paper. Therefore, several criteria that were created during the process of this review present slightly adjusted variations of the original. Yet, other criteria do not directly relate to any of the initiatives at all, although the inspiration for those can be drawn from them. For instance, the assessment of key performance indicators of the Governance Institute of Australia, did not provide any concrete criteria, but led to the creation of criteria 13 and 19 (see Appendix 1)**.** In the end, the RobecoSAM Survey and the World Economic Forum's Global Competitiveness Report aided in the creation of each 6 criteria, making them the most impactful initiatives for this assessment framework, followed by the World Bank Governance Indicators which contributed directly to the creation of 4 criteria. Beyond this, the assessment of the Institute of Director and Equitable Origin 100 provided 3 criteria each and the other mentioned initiatives up to 2. From the set of 38 criteria that were eventually used in the final stage (See Figure 7) of the creation of the assessment framework, 26 can be linked directly to the examined initiatives. It is important to note that during the process of the creation and selection of criteria and indicators (See Figure 7), several decisions and choices led to cuts and refinement of the set of criteria that is presented below.

Not only was there a process of selecting, examining and eventually deciding on suitable initiatives (Stage 1, Figure 7), where a set of 27 initiatives and scientific literature was downsized, based on a brief review of the categories and sub-categories, to 12 potentially useful initiatives (see Table 2), but also has there been a process of closely examining all criteria of these 12 initiatives. About 200 criteria from these initiatives were considered to be of potential use for the evaluation (Stage 2, Figure 7) and were thus checked for their applicability to the standard system context. Out of those, about 140 criteria were reformulated (if necessary) (Stage 3, Figure 7) and put into the appropriate sub-category of the first draft of the assessment framework. However, after yet another close examination aimed at avoiding that certain aspects are counted more than once, checking whether the criteria are relevant in reality and due to a need to keep the criteria set manageable further cuts led to stage 4 (see Figure 7), where a set of 65 criteria was compiled.

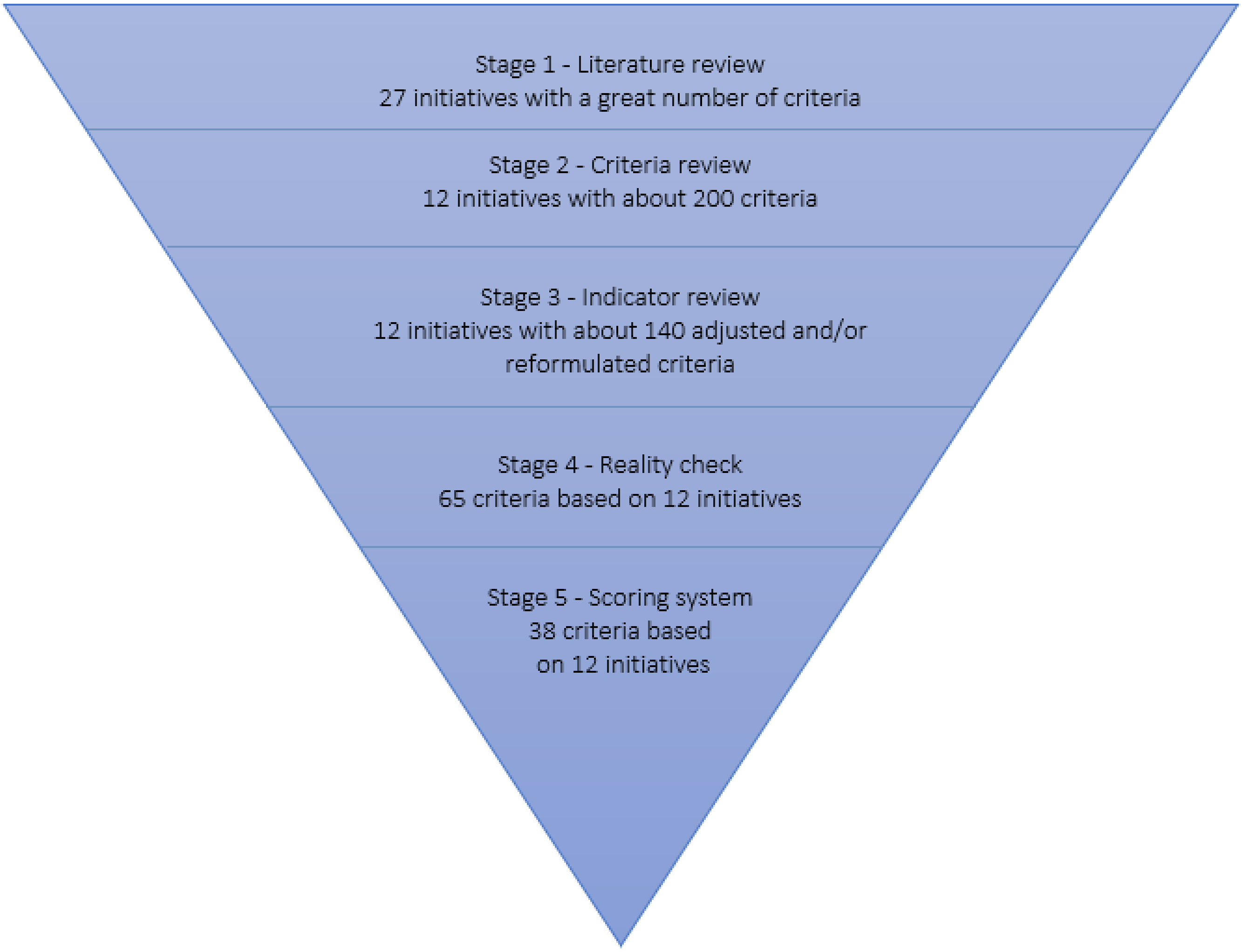

Figure 7 – Stages of the assessment framework creation process

Finally, the fifth stage (see Figure 7) was reached by creating the exact indicators and scoring system, leading to a total of 38 criteria. The 17 criteria that were cut out at this point would require an intensive amount of study or were not quantifiable in a meaningful manner. For example, ‘Years with the current auditor’ (originally a criterion of the ‘Internal accountability’ sub-category in Stage 4) would be too superficial as a longer duration does not necessarily indicate the maintenance of a fruitful relationship. Throughout this process sixteen different versions were created in Excel, each of which was an improvement upon the former until a first version of the assessment framework was finalised and finally applied to the FSC, PEFC, UTZ/RA and GCP. The assessments, as described in the methodology section (Chapter 3) ought to go through two assessment rounds, including interviews, but after the finalisation of the first assessment rounds, due to time constraints, had to be limited to communication via e-mail. The standard organisations were requested to fill in missing information, providing additional detail and comments on the framework. From the four examined standards, only the FSC, through Marion Karmann, responded and helped with the partial completion of the FSC assessment. The following chapter describes the current version of the assessment framework with more detail and is followed by the assessments.

4.5 Assessment Framework for Standard Systems

The first complete version of the assessment framework consists of five categories which essentially are the five principles of UNDP good governance concept and therefore are named precisely after those principles. Within these categories, two to four sub-categories with a varying number of criteria are used to evaluate good governance in voluntary sustainability standard systems. The sub-categories are based on the evaluation of good governance in standard systems (chapter 4.2) and explained in more detail below. Within those 13 sub-categories (see Table 3), 38 criteria can be found.

Category	Weight	Sub-category	Weight
Legitimacy and Voice	20%	Appropriate basis	33%
		General quality	33%
		Answerability - Formal involvement	33%
Direction	20%	Goals and targets	50%
		Theory of change	50%
Performance	20%	Progress towards goals and targets	50%
		Responsiveness - Actual involvement	50%
Accountability and Transparency	20%	Transparency - Availability of information	25%
		Pre-interview information availability	25%
		Internal accountability	25%
		External accountability	25%
Fairness	20%	Equity*	50%
		Rule of law	50%

Table 3 – Weighting system of the assessment framework (*Equity sub-criteria contribute equally to each criterion)

Table 3 visualises the weighting system of the assessment framework, showing that each of the sub-categories has an equal impact towards their category, which all have the same weight towards the final score. An equal weighting system was chosen as there is, at this point, no reason to assume that one category should be more important than another. Aspects like, for instance, the significance of stakeholders has been worked into the framework by creating more criteria and sub-categories that deal with those, effectively granting those aspects more weight. Leaning on the Institute of Directors' (2017) Good Governance initiative, every criterion will be awarded a score between 1 and 0 together making up the score of the sub-categories, which in turn lead to the scores for the five categories and the final score. Most of the criteria have ordinal indicators, awarding a score of 1, 0.5 or 0. It must be noted that, at this point, it was difficult to judge certain aspects without further testing of the framework, hence some indicators may need a serious review. For instance, criterion 13 'Time frame of the vision', awards a score of 1 if the vision covers more than the next 10 years, a score of 0.5 for less than 10 years and a score of 0 if there is no clear statement on the vision. It may be that this time frame is inappropriate, causing the score of 0.5 to be irrelevant or perhaps making the requirements to reach the score of 1 too ambitious and unattainable.

The equity sub-category is slightly different from the rest of the framework, as it is the only sub-category that has several sub-criteria that all have an equal weight towards their criteria. A reason for this is that the review of initiatives and their subsequent criteria provided an extraordinary quantity of useful objective criteria with comparatively simple data requirements.

Beyond this, another special thing about this framework is the existence of some criteria being connected to one another, where the former must be answered in order to judge the latter. These are criteria 9, 10 and 21 (stakeholder coverage and attendance) as well as 13 and 19 (self-assessment). As there is a risk that these connections grant the first of each of these criteria more weight, since a failure in the first would mean failure in the latter, the latter are rated as 'Not applicable' if the former is scored 0. More detailed information on how each criterion is assessed can be found in the blank version of the assessment framework and the user guide provided in the appendix (Appendix 1 and 2).

'Legitimacy and voice' (see Appendix 1) is divided into three sub-categories being the 'appropriate basis', 'general quality' and 'answerability'. The first sub-category examines whether or not the system provides the necessary basis for the development of a high-quality standard, which includes having qualified people, functioning structures and routines et cetera. Building on this, the second sub-category, 'general quality', broadly examines factors that signal a high quality, which can be divided into novelty and innovation, moral authority and necessity and usefulness of the standard. Such a high quality is essential for the maintenance of the authority of the standard. Lastly, the third sub-category, determines whether a wide range of stakeholders are, *formally* involved as a part of the system through mechanisms, policies and so on. Due to a great importance of stakeholders, there are two more sub-categories in other parts of the assessment framework which examine other aspects of stakeholder involvement.

The second category, 'direction' (see Appendix 1), emphasises the need for a strategic vision and is therefore split into the categories 'goals and targets' and 'theory of change'. For obvious reasons it is fundamental to have and strive for certain 'goals and targets' to create a sense of direction in the system. Beyond this a theory of change is useful to streamline efforts and maintain a consistent approach. While its aspects can be of a varying quality, it is something that cannot be analysed within the scope of this research. However, criterion 14 checks for the existence of the theory of change building blocks (as explained by Komives (2014)), while criterion 15 is used to examine the scientific basis of these.

'Performance' (see Appendix 1) is determined by the sub-categories 'progress towards goals and targets' and 'responsiveness – actual involvement'. The former category is mainly about the effectiveness and efficiency of the system; the ability to reach their targets, the efficiency of their bureaucratic activities etc. but also about checking their own progress and actively trying to improve this. 'Responsiveness' is the second sub-category that focuses on stakeholders and complementary to the sub-category found in the legitimacy category, it is supposed to examine if stakeholders actually get actively involved.

For the ‘accountability and transparency’ category (see Appendix 1), three sub-categories are used. ‘Transparency - availability of information’ in an addition to the result of the first round of the assessment (criterion 24. public availability of information), consisting mainly of an examination of the availability of vital information. It should be noted that the ‘availability’ column (Column D) is used to check how much information is publicly available and where an interview could reveal missing/additional information. This will factor into the scoring of criterion 24, which, theoretically, has the most weight any single criterion in the assessment framework holds (see Table 3). Information obtained in the second assessment round was still counted as ‘information available pre-communication’ if it was merely missed during the first round but was publicly available. Beyond this, the second part of this category is made up of the ‘internal-‘ and ‘external accountability’. Internal accountability deals with the relationships between accreditors, auditors and the standards organisation as well as internal procedures and limitations on the decisionmakers, while external accountability presents the final sub-category for the measurement of stakeholder involvement, at this point being the determination of whether stakeholders actually have any power to push for their own agenda.

In the final category, fairness (see Appendix 1), the two theoretical aspects of the UNDP good governance definition have been transferred rather directly to the assessment framework. The ‘equity’ sub-category examines whether employees have fair opportunities, while the ‘rule of law’ sub-category mainly focuses on adherence to the law.

4.6 Application of the assessment framework

During this first review round, where the PEFC and the FSC were examined, some aspects of the framework already began to stand out and were adjusted before finalising the first assessment round of these two organisations. These changes were mainly related to changes in formulation and precision. Since standard setting organisations appear to be relatively small regarding the number of employees, not much information on the fairness aspects was available. Despite a claim to be transparent, there was little information on various work-related policies, worker’s organisations, dialogue amongst one another, employee turnover rates et cetera. However, the following chapters describe the first assessment rounds aka. pre-communication assessments. The assessment of each standard is presented with a spider web figure for an improved visualisation. For each category, extraordinary factors are explained as there have been several irregularities caused by a lack of data or an unclarity of crucial information.

4.6.1 Assessment of the Forest Stewardship Council

The FSC has been examined as the first of four different standard systems. After examining the website and relevant documents, 31 (83.78%) of the 37 (excluding the transparency criterion) criteria could be answered to some degree, allowing for the development of an initial score.

Based on data found during an unassisted search period of one week, followed by an evaluation of this information, the scores that can be seen in Figure 8 were developed. As one might expect the FSC to be a prime example of good governance, the direction and fairness scores indeed seem to hint at surprising shortcomings that can be explained when looking at the individual criteria of these categories. However, these two categories and certain individual criteria have turned out to be affected by more overarching issues, which will be described in Chapter 5.1.

The FSC reached a score of 0.917 in the first category with mostly perfect scores, only marred by a score of 0.25 for their ‘research activities’ (criterion 4). This score was based on it being rather unclear to what extent the FSC gets involved in those research activities and could potentially change after communicating with experts from the FSC. The direction score was a 0.583, making this the highest of the four direction scores. Indeed, it appears that the FSC has rather clear goals and targets as well as a well-developed theory of change. However, mainly criteria 12 and 15 dragged the score down and are prone to change in a future version of the framework. At a score of 0.833 for the performance category, it can be said that the FSC performs very well, the only criterion that could be improved upon being criterion 19, the reporting on their self-assessment method. Despite lacking information on one of the criteria in the accountability and transparency category, the FSC certainly deserved its perfect score of 1 in this category, as the standard setting organisation cooperates with other organisations within their system to an extent that was not noted as such in the other three schemes. Lastly, the fairness category was scored 0.542, which was mainly affected by more general issues that are explained in Chapter 5.1. At the end of the first round, the total score of the FSC was a 0.775 and therefore almost the highest of the four assessments

With the help of Marion Karmann, who is active in the M&E program at the FSC, the legitimacy and voice, direction and performance categories could be answered more thoroughly. Although this would not suffice to claim a full second assessment round has been completed, it provided a great addition to the assessment of the FSC. Now a total of 34 criteria could be scored, although some criteria in, especially in the fairness category, would be expected to strongly benefit from more expert feedback. Additionally, the additional sources allowed the adjustment of a few scores. While only little has changed in the legitimacy and voice as well as the performance categories, the direction category was impacted to the greatest extent. The self-assessment (criterion 13) appeared to be more detailed than expected and thus has been awarded a full score. This raised the direction score from 0.583 to 0.667 and the total score to 0.789.

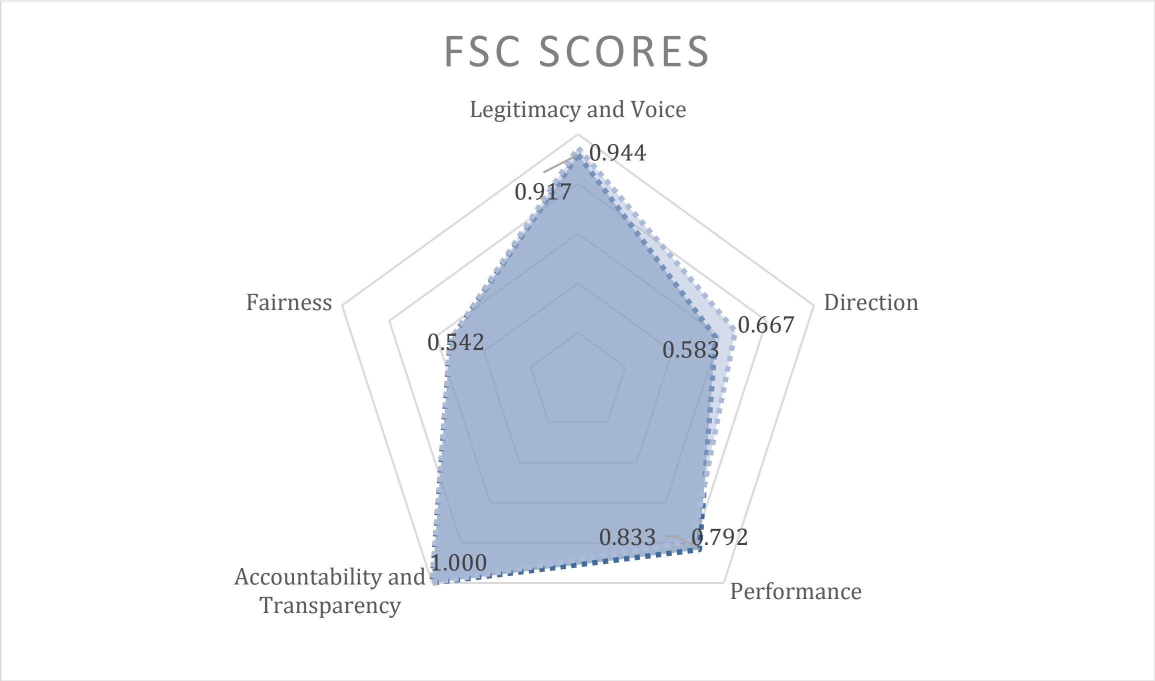

Figure 8 – FSC scores pre- and post-communication

4.6.2 Assessment of the Programme for the Endorsement of Forest Certification

For the PEFC, sufficient information to answer 33 (89%) of the criteria could be found. Just like for the assessment of the FSC, a full week was spent trying to fill the assessment framework with information from publicly available sources. Regarding the legitimacy and voice category, only one criterion stood apart from an otherwise admirable performance, leading to a score of 0.944 in this category (see Figure 9). The members of the PEFC Council appear to predominantly originate from the global north (68.75%), leading to a decreased score for criterion 9. In the direction category, the PEFC scored a 0.292 which is relatively low and might be related to the lacking availability of a document that described their strategy. Merely a summary of said strategy is published, based on which many of the criteria in this category were judged.

Despite certifying the highest volume of timber, the performance score of the PEFC (0.750) was lower than that of the FSC. A main factor here was what seems to be a lack of support programmes directly aimed at producers (criterion 17), since most projects of the PEFC focus on supporting national systems or providing group certification. Although group certification is very useful for producers, it was not considered as a part of this assessment framework and could potentially be included in the future. The accountability and transparency score of 0.958 was only negatively impacted by criterion 28. While the PEFC appears to have refined mechanism to check its auditors, this only really happens during the first time an auditor seeks accreditation and whenever the standard itself is adjusted. Aside of the general issues in the fairness category, the PEFC's score of 0.5 was affected by a low score for criterion 37 (Support for a fair and just environment) for the same reasons it scored low in criterion 17 (described above).

The PEFC final score for the first round of the assessment was a 0.689.

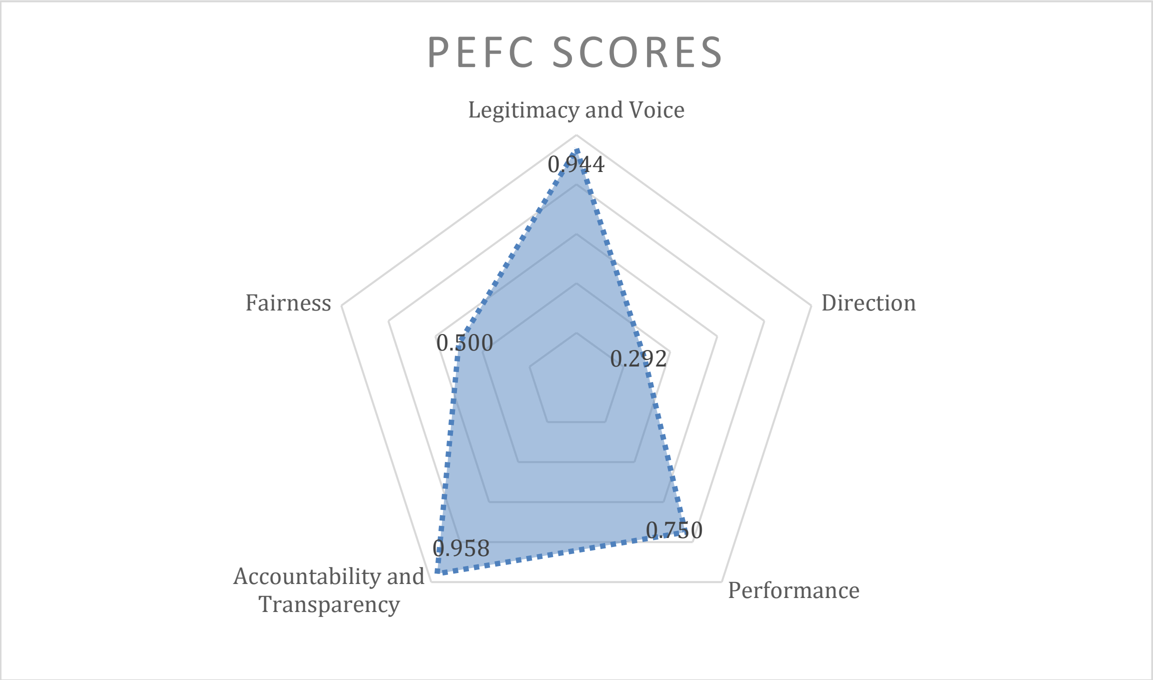

Figure 9 – PEFC scores pre-communication

4.6.3 Assessment of UTZ Certified / Rainforest Alliance

The results of the UTZ/RA were very surprising as the assessment of this standard went relatively smoothly, finishing the collection of data within four days with a total of 28 (75.68%) of the criteria sufficiently answered. It might be related to the ten insufficiently answered criteria which effectively led to certain criteria having an unusually large impact during this first assessment round. However, some of these criteria indeed appeared to not be applicable as there seemed to be no meetings where all stakeholders are formally invited to discuss certain aspects of the scheme. While stakeholders do participate in the standard development procedure, this would not qualify as the kind of meeting/membership described in criteria 9 and 10.
The legitimacy and voice score of 0.806 was this 'low' because of criterion 7 (Relative added value of the standard) (see Figure 10). Unlike all other schemes, UTZ/RA seems to shy off from trying to clearly point out its strengths and remains quite neutral and objective in the presentation of its benefits, without ever trying to claim being the best at a certain aspect or boast about anything else. Although there certainly are aspects to the UTZ/RA where it clearly sets itself apart from similar schemes, it was supposed, for the purpose of this criterion, to point towards those differences on its own and therefore scored rather low. Similar to the PEFC, the direction category score of the UTZ/RA suffered from a lack of a document that provided details on their strategy, resulting in a score of 0.5. Regarding the performance score, certain 'not applicable' criteria demonstrated a serious impact. Having only one valid criterion, which scored a 0 (criterion 20; Consistent formal dialogue with stakeholders), meant that half of the performance score was affected by a single criterion, which in this case, could also not be judged properly without

feedback from experts of the UTZ. If this single criterion would achieve a score of ‘1’ after the communication, it would raise the total performance score from a 0.472 to a 0.972. Accountability and transparency reached a score of 0.875, only being by affected by a lack of information on some form of auditor control process (criterion 28). The UTZ/RA explains its auditor approval procedure and mentions that certification bodies have to ensure that they update their procedures whenever the standard is changed. Lastly, the fairness category scored 0.306 and had exactly the same problem that was found in the performance category. In the ‘rule of law’ sub-category information for only one criterion (criterion 36) was found, which impacted the total score. Would this single criterion be scored with a 1, the entire category would rise to a score of 0.806. Overall the UTZ/RA assessment definitely required additional input from experts who are more knowledgeable on the functioning of this scheme to properly assess it. As for the first round of this assessment the UTZ/RA reached a total score of 0.592 mainly because of the lack of data, which caused the two mentioned criteria to have an unproportionate impact.

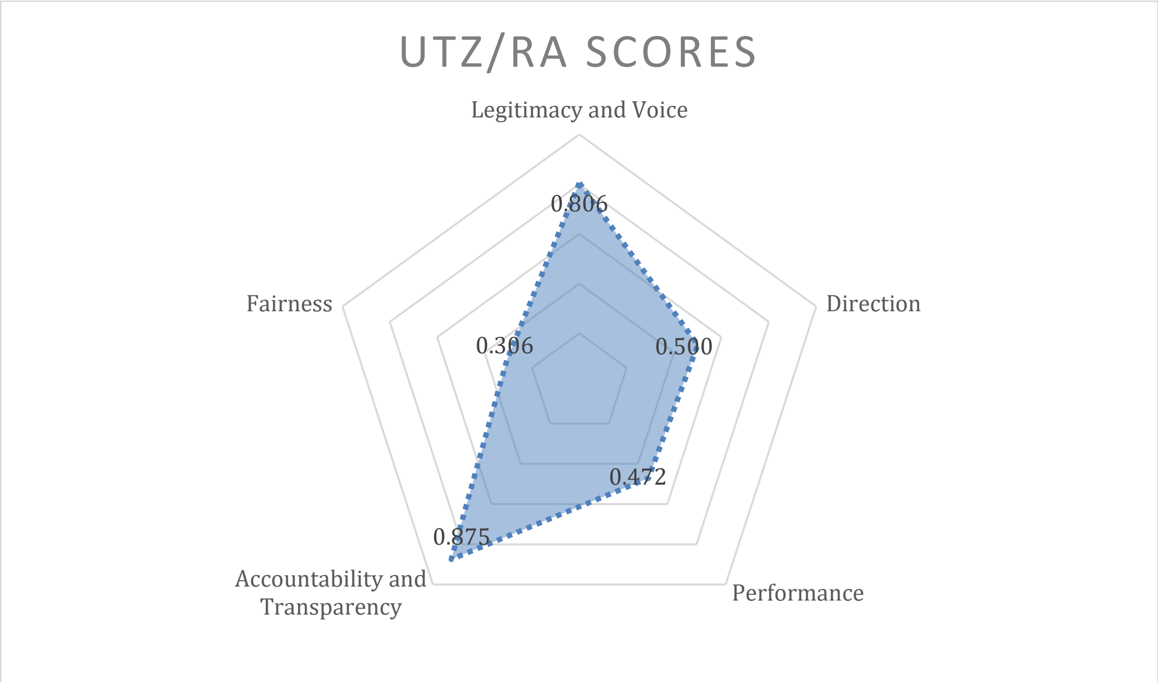

Figure 10 – UTZ/RA scores pre -communication

4.6.4 Assessment of the Global Coffee Platform

The GCP reached almost the same initial total score as the FSC and information collection required about as much time as it did for the FSC and PEFC. Eventually 34 (91.892%) of the 37 (excluding the special transparency criterion) criteria could be scored based on publicly available information, which also was the highest number of criteria that were answered in any of the first assessment rounds. Most impactful in the legitimacy and voice category (0.889 score) was criterion 2 (accreditation authority), which had a slightly worse score because it seemed that the

GCP's accreditors are not fully independent third parties (see Figure 11). The score of 0.375 for the direction category was induced by, once again, a lack of a clear strategy. While there appears to be the 'Vision2020' project, there did not seem to be much information available on it (e.g. the website "http://vision2020.coffee/", linked by the GCP, is unavailable). Regarding the performance score the GCP reached a perfect score of 1, although it was difficult to judge some aspects like criterion 18 (relative impact of the standard) as it examined the volume of certified products, which is something more easily achieved by the GCP as it is a 'first step' towards sustainable coffee and, logically, easier to obtain than comparable schemes. In the accountability and transparency category (scored 0.906) only criterion 27 (accountability of the auditor) and 28 (accountability of the auditor) did not reach a full score. The accreditor seems to only be contacted when the standard is changed, and auditors go through a re-verification process every three years. The GCP was the only organisation that scored relatively good in the fairness category, with a score of 0.689, which was largely based on a good score in the rule of law sub-category. In the end, the total score after the first assessment round was a 0.774.

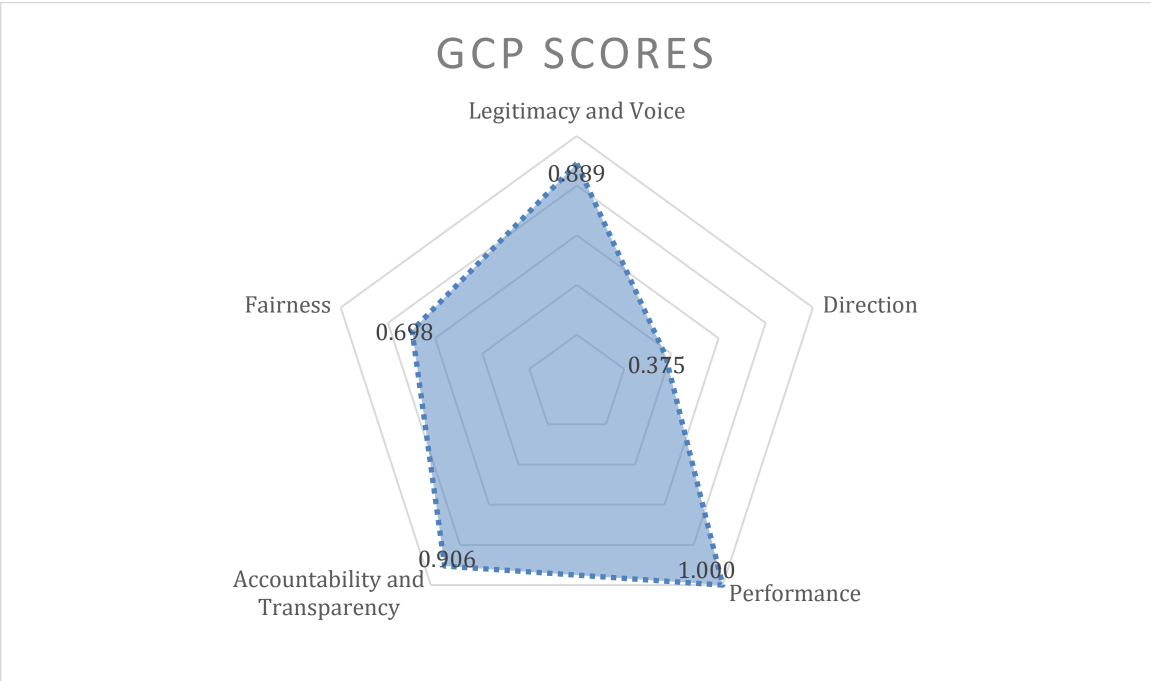

Figure 11 – GCP scores pre-communication

In the end the FSC and GCP came out on top during the first round of the assessment with scores of 0.775 and 0.774 respectively (see Figure 12). Followed by that was the PEFC with a score of 0.689 and the UTZ/RA with 0.592. The issues with those were explained above, but the rather high scores do imply that large and well-established standards do, consciously or not, make an attempt at conducting good governance. Especially because it should be expected that, after a properly conducted second assessment round, these scores should be even higher. There is no set threshold at which it can be said that one organisation conducts 'good governance' while another does not. These scores should be interpreted as a certain degree of good governance that has been achieved by the standard system.

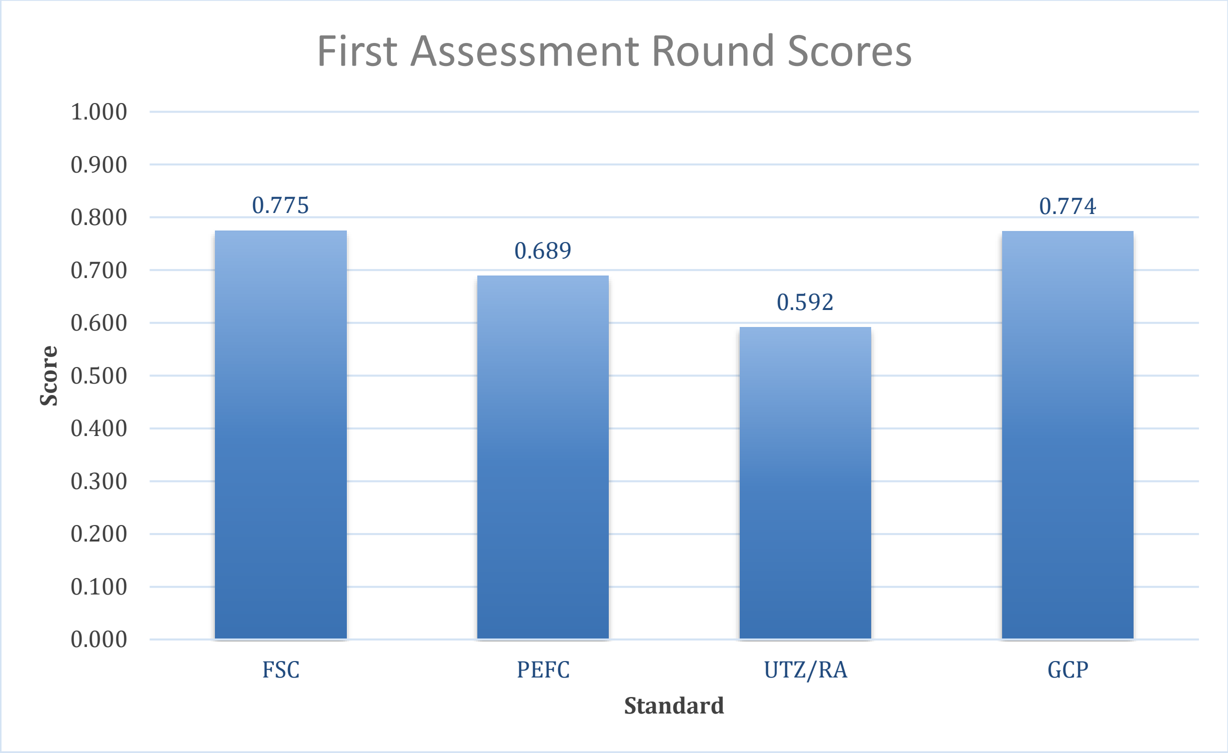

Figure 12 – Final scores of the first assessment round

5. Evaluation

5.1 Reflection and Optimisation of the Framework

After finalising the assessments of the PEFC, FSC, GCP and UTZ/RA a number of criteria and indicators require adjustments that may not be done straight away, depending on the issues that have been uncovered by this application and the feedback from the standard setting organisations. This chapter first examines aspects of individual criteria and then proceeds to discuss broader issues, based on the first phase of the assessment (based on public sources). Afterwards additions will be described that were enabled through the feedback of experts from the standard setting organisations. In the blank excel sheet, the 'sources' column contains more detailed remarks that are described in this chapter (see Appendix 1).
During the first round of the assessments it became more visible that even the standard setting organisations of rather large standards have far fewer employees than any typical private sector actor that has been examined by the initiatives that were used as a basis for the assessment framework. Accordingly, certain features were just not required or did not exist.

Predominantly the fairness category seemed to be affected by this issue as it assumes that the employees could have a far greater impact than they actually have on the standard setting organisations. In all four cases, publicly available information on workplace policies (criteria 5, 31.1, 31.2, 31.3) was rather scarce, likely because conflicts are not as common as in larger private organisations. Although, these policies are beneficial, they may not be as relevant if there is little conflict and a lot of informal dialogue. Criterion 33.1 (Dialogue between workers and management) also was affected by the same issue as informal dialogue may very well suffice for the workers themselves to express their concerns to the management, without the need for a committee or a worker's organisation to play this part. An entirely different issue occurred with the criteria for equality. Since both of these criteria (32.1 and 32.2) focused on a balance between male and female individuals, it could lead to false conclusions as these will be limited to gender equality and beyond this, an imbalance does not necessarily mean that there is inequality. The other initiatives from the literature review that examined equality had additional criteria to assess equality in broader terms. Lastly, the employee turnover rate (criterion 34.1) was on the one hand difficult to obtain, but on the other hand might have had incorrect indicators. Sources conflict over what turnover rate is ideal, since a low turnover rate may indicate a harmful stagnancy or stability of a well-functioning environment.

Another aspect that was crucial during the first round of the assessment was the relevance of certain sources of information that would impact multiple criteria at once and prove rather problematic if one of these sources was not formulated or available. For instance, a complete list of activities (mainly projects) and details on these could be used to deal with a number of diverse criteria such as 'Support for a fair and just environment' (criterion 37), 'Support programmes for producers' (criterion 17) and 'Research activity' (criterion 4). While the FSC and PEFC provided a rather clear overview on their projects, the UTZ/RA and GCP assessments demanded additional investigation to find information about relevant activities.

Beyond this, the vision is often displayed with a varying degree of detail which did have a dramatic impact on the direction category. A document that describes the strategy and/or future planning of the standard is useful for the 'Precision of the vision', 'Time frame of the strategy', 'Theory of change' and 'Empirical validation of the theory of change' (criteria 11, 12, 14 and 15). This means that a lack thereof created a very low score as demonstrated in the first assessment round of the PEFC. Other comparable impactful sources of information were the detail on members and stakeholders (criteria 9 and 10), the complaint/incident reporting mechanism (criterion 29), stakeholder interaction (criterion 8) and the self-assessment method (criteria 13 and 19).

When individually looking at the criteria, especially during the first assessment round, the overarching issue has been a lack of data. While many criteria could only be partially answered, it was not possible to find sufficient information for the seven following criteria in any of the four first round assessments; 'Share of academics among the employees in the standard setting organisation', 'Evaluation of ethics', 'Empirical validation of the theory of change aspects', 'Bureaucratic delays', 'Employee protection/whistle-blower policy', 'Employment policy' and 'Support for a fair and just environment' (Criteria 1, 5, 15, 16, 31.1, 31.2, 37). All of these criteria were deemed necessary for good governance and supposedly applicable to voluntary standard systems during the development of the framework. While for most of these it can just be concluded that information simply is not readily available, it would seem that an evaluation of their own ethics (criterion 5) is something that does not occur to standard's organisations as they usually promote ethical products and/or production methods. However, promoting ethical practices elsewhere does not necessarily extinguish the presence of unethical practices within their own organisation(s).

Another complication were indicators, mostly scored with a 0.5, that described a situation where a claim is made but no evidence would be provided to support this claim. This kind of indicator was created for criteria 5, 8, 31.1, 31.2, 31.3 with the purpose of creating fewer absolute scores (1 or 0) but did turn out to be inapplicable as evidence could always be found.

In a contrast to this, there appear to be yet again other criteria where achieving the full score was rather easy, meaning that providing additional detail should perhaps be scored higher by creating partial scores. Among those criteria were the ‘recognition of the issue focused on by the standard’ (criterion 6), ‘standard criteria, indicators, rating systems etc.’ (criterion 23) and ‘exchangeability of stakeholder representatives’ (criterion 30). Similarly, there were several criteria where a choice had to be made with regards to the (time) frames for certain scores, without having elaborate background information on these numbers for standard systems. This effectively led some scores to be unachievable or too easy to obtain. Criterion 12 (time frame of the strategy) required the strategy to be planned for the next ten years, which appeared to be inappropriate considering that even the FSC only plans for six-year cycles. The same occurred for the division of executive power (criterion 25) which was scored as a 1 if the power was split between 3 or more persons, however, standards organisations have boards with ten or so member who are responsible for day-to-day activities and decision-making. Criteria 1 and 16 may be subject to the same issue but could not really be assessed after the first round.

There were some broader aspects of standard setting organisations that did affect the usefulness of the framework, the greatest of which was the division of legal entities and subsidiary organisations. The PEFC and FSC are very large schemes that work with national units that overtake diverse parts of the process, which made it difficult to judge some criteria as they may be dealt with in a satisfactory manner in one national unit, while another one might be unable to fulfil this. Even when trying to focus exclusively on the international unit, this may cause certain aspects to seemingly not exists, while those could simply be part of the tasks that subsidiaries overtake. Furthermore, when details of the employees of the standard setting organisations were examined, it could be very difficult to determine which employees were important to the criterion. Other extraordinary features, such as those created by the recent merger of UTZ Certified and the Rainforest Alliance or the transformation of the 4C Association into the Global Coffee Platform also complicated the scoring of certain criteria and led to unclarity. Such features must be accounted for unless the assessment framework will further develop into a direction that only examines rather unspecific and broad aspects.

Although a lot of effort was put into downsizing the number of criteria of the framework, it appears that it might be beneficial to expand it again as that can resolve a couple of problems. With the possibility, induced by a lack of data or inapplicable criteria, that an incomplete set would be used to evaluate the standard systems, there comes the risk of a single criterion gaining far more impact than it ought to have. This was demonstrated in the first assessment round of the UTZ/RA’s performance and fairness scores. Expanding to a minimum of three criteria per sub-category would decrease this risk and also allow for a better coverage of certain aspects, such as equality (criteria 31.1 and 31.2), which was essentially only looking at gender equality. Despite other original plans, the application of the framework was effectively conducted by an independent third party only with raw information as input from the standards organisation but changing this approach might be valuable. Experiencing that the expert from the FSC was able, with a lot of clarification, to answer most of the questions, it might be possible to adjust this design to allow standards organisations to fill it themselves as a self-assessment tool, much like the Global Compact Self-Assessment Tool. However, the 'User guide' would require a lot more detail for this purpose as it currently is not as expansive.

5.2 Limitations of the Research

Looking back at the research as a whole, a number of drawbacks could be found. One major limitation was the coverage of the literature review (Chapter 4.3) which could only check so many initiatives and might have made the creation of the assessment framework a lot easier if more directly applicable criteria were found. For instance, the MSI Evaluation Tool was designed specifically for multi-stakeholder initiatives but was published just recently (in November 2017) and thus was missed during the literature review phase as there is not really much research based on this new tool. Although this tool does not exactly examine good governance, it might have produced useful criteria for this assessment. Besides this, most criteria of the framework had to be adjusted to a great extent to be suitable to assess voluntary sustainability standards. Backtracking these criteria would be useful as some of them might have been adjusted to an extent that would invalidate the claim of it being based on a certain source.

Besides this, it was not expected that the standard organisations are so far more diverse than they appeared to be based on the literature. Unique aspects must be covered without moving towards a model that has to be tailored to every single case. Hence, the functioning of standards chapter could be more detailed (Chapter 4.1) and potentially divide standards into certain categories that would be assessed slightly differently. The greatest limitation to this research, as so often appears to be the case, was the time limit. Seeing that some initiatives needed years to complete (e.g. the MSI Evaluation Tool took seven years to be finalised), it did not come as a surprise that this good governance assessment framework could not be completed. Affected by this lack of time, besides the mentioned literature review, were mainly the assessment phase and consequently the finalisation of a new version of the framework. Having initially planned to conduct interviews with experts from the standards organisations to receive elaborate answers to every single criterion and eventually ending up with being limited to communication via e-mail, requesting the experts to fill out the form, seriously inhibited the completeness of the research. Especially because only the FSC was able to provide a (partial) response in time. Arguably, the communication could have started a bit earlier, potentially after the fourth stage of the framework (Figure 7), but that would have meant that obvious and easily fixable flaws in the framework would persist. Although it was indeed planned to step into contact much earlier than this, the task of creating the assessment framework was underestimated.

6. Conclusion

With the main research question being; “*How can good governance of voluntary sustainability standard systems be analysed?”,* this research tried to develop a method to analyse good governance in standard systems in a, preferably, objective and comparative manner. Indeed, the developed assessment framework is on the correct path towards being able to provide such a method. The analysis showed that it is possible to evaluate good governance in standard systems, despite rather few attempts at doing so and most of those being tailored towards a single standard. It is true that the diversity among standard systems is great and every standard is unique in multiple ways, yet they do share certain core features based on which a comparison is possible when enough information is available. This comparative tool, when perfected, hopefully provides consumers with the ability to differentiate more easily between standards and helps standard organisations to identify strengths and weakness, allowing them to improve upon such aspects. Eventually such improvements should lead to better governance and therefore a greater effectiveness of the standards and their impact on our path towards achieving sustainability development, which after all is one of the greatest tasks of our time.

However, the current version of the framework is incomplete and requires more work. Future efforts should be directed at improving aspects of the assessment framework that remain lacklustre. While background information should be improved, the focus of a future study should lie in cooperation with experts from the field through more detailed trial assessments of the framework as was already initiated, but not completed, in this study. These additional and more elaborate assessments should be based on a version of the framework that has been improved through the findings of the reflection of this study (Chapter 5.1).

References

Aguilera, R. V., & Cuervo-Cazurra, A. (2004). Codes of good governance worldwide: What is the trigger? *Organization Studies*, *25*(3), 415-443.

Arat, Y. (1991). Politics and big business: Janus-faced link to the state. *Strong state and economic interest groups: The post-1980 Turkish experience*, 135-147.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The database of political institutions. *the world bank economic review*, *15*(1), 165-176.

Bernstein, S., & Cashore, B. (2007). Can non-state global governance be legitimate? An analytical framework. *Regulation & Governance*, *1*(4), 347-371.

Besançon, M. (2003). Good governance rankings. *The art of Measurement. World Peace Foundation. WPF Reports*, (36).

Bitzer, V., Francken, M., & Glasbergen, P. (2008). Intersectoral partnerships for a sustainable coffee chain: Really addressing sustainability or just picking (coffee) cherries? *Global Environmental Change*, *18*(2), 271-284.

Blackman, A., & Rivera, J. (2011). Producer-level benefits of sustainability certification. *Conservation Biology*, *25*(6), 1176-1185.

Bovaird, T., & Löffler, E. (2003). Evaluating the quality of public governance: indicators, models and methodologies. *International Review of Administrative Sciences*, *69*(3), 313-328.

Castka, P., Corbett, C. (2016). Adoption and diffusion of environmental and social standards: The effect of stringency, governance, and media coverage. *International Journal of Operations & Production Management*, *36*(11), 1504-1529.

Clague, C., Keefer, P., Knack, S., & Olson, M. (1999). Contract-intensive money: contract enforcement, property rights, and economic performance. *Journal of economic growth*, *4*(2), 185-211.

Coffee Assurance Services (2018). Information on the Coffee Assurance Services. Retrieved from https://www.cas-veri.com/

Cruz, C., Keefer, P., & Scartascini, C. (2016). The Database of Political Institutions 2015: Codebook Cesi Cruz. Retrieved from https://www.iadb.org/en/research-and-data/publication-details-0?pub_id=IDB-DB-121

De Felice, D. (2015). Business and human rights indicators to measure the corporate responsibility to respect: Challenges and opportunities. *Human Rights Quarterly*, *37*(2), 511-555.

Denny, D. M. T. (2017). Voluntary Sustainability Standards. *International Economic Law and the Environment.*

Dingwerth, K., & Pattberg, P. (2009). World politics and organizational fields: The case of transnational sustainability governance. *European Journal of International Relations*, *15*(4), 707-743.
Doornbos, M. (2001). 'Good governance': The rise and decline of a policy metaphor? *Journal of Development studies*, *37*(6), 93-108.

Edgar, L., Marshall, C., & Bassett, M. (2006). *Partnerships: Putting good governance principles in practice*. Ottawa, Ontario, Canada: Institute on Governance.

Equitable Origin (2012). EO100™ Standard – For Responsible Energy Development. Retrieved from https://www.equitableorigin.org/

Esty, D. C. (2005). Good Governance at the Supranational Scale: Globalizing Administrative Law. *Yale lj*, *115*, 1490.

Feng, Y., Kugler, J., & Zak, P. J. (2000). The politics of fertility and economic development. *International Studies Quarterly*, *44*(4), 667-693.

FSC (2018). Information on the Forest Stewardship Council. Retrieved from https://ic.fsc.org/en

GCP (2018). Information on the Global Coffee Platform. Retrieved from http://www.globalcoffeeplatform.org/

Graham, J., Amos, B., & Plumptre, T. (2003). Principles for good governance in the 21st century. *Policy brief*, *15*(6).

Henisz, W. J. (2000). The institutional environment for economic growth. *Economics & Politics*, *12*(1), 1-31.

Hood, C., Dixon, R., & Beeston, C. (2008). Rating the rankings: Assessing international rankings of public service performance. *International Public Management Journal*, *11*(3), 298-328.

Institute of Directors (2017). *The 2017 Good Governance Report – The Great Governance Debate Continued.* Retrieved from https://www.iod.com/news/news/articles/The-2017-Good-Governance-Report

International Accreditation Forum (n.d.). Why use an accredited certification body? Retrieved from http://www.iaf.nu/upFiles/IAF_Why_use_accredited_CB_0112.pdf

International Trade Centre (2018). Sustainability Map, Your Roadmap to Sustainable Consumption, Production and Trade. Retrieved from https://sustainabilitymap.org/home

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). Governance Matters VIII: Aggregate and Individual Governance Indicators 1996-2008. *World Bank Policy Research Working Paper No. 4978, 1-105.*

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, *3*(2), 220-246.

Knoll, M., & Zloczysti, P. (2012). The good governance indicators of the millennium challenge account: How many dimensions are really being measured? *World Development*, *40*(5), 900-915.

Komives, K., & Jackson, A. (2014). Introduction to voluntary sustainability standard systems. In *Voluntary Standard Systems* (pp. 3-19). Springer Berlin Heidelberg.

Komives, K. (2014). Developing a Theory of Change: The foundation of a robust M&E system [Webinar]. Retrieved from https://www.isealalliance.org/get-involved/resources/developing-theory-change-foundation-robust-me-system

Marshall, M., Jaggers, K., & Gurr, T. (2017). Political regime characteristics and transitions, 1800-2016, Polity IV Project.

Marshall, M. G., & Jaggers, K. (2002). Polity IV project: Political regime characteristics and transitions, 1800-2002.

Marx, A., & Wouters, J. (2014). Competition and cooperation in the market of voluntary sustainability standards. *Browser Download This Paper*.

Mkandawire, T. (2007). 'Good governance': the itinerary of an idea. *Development in Practice*, *17*(4-5), 679-681.

PEFC (2018). Information on the Programme for the Endorsement of Forest Certification. Retrieved from https://www.pefc.org/

Phillips, J., Hailwood, E., & Brooks, A. (2016). Sovereignty, the 'resource curse'and the limits of good governance: a political economy of oil in Ghana. *Review of African Political Economy*, *43*(147), 26-42.

Ponte, S., & Cheyns, E. (2013). Voluntary standards, expert knowledge and the governance of sustainability networks. *Global Networks*, *13*(4), 459-477.

Reinecke, J., Manning, S., & Von Hagen, O. (2012). The emergence of a standards market: Multiplicity of sustainability standards in the global coffee industry. *Organization Studies*, *33*(5-6), 791-814.

RobecoSAM (n.d.). Corporate Sustainability Assessment – Sample questionnaire. Retrieved from http://www.robecosam.com/en/sustainability-insights/about-sustainability/corporate-sustainability-assessment/sample-questionnaire.jsp

Roy, K. C., & Tisdell, C. A. (1998). Good governance in sustainable development: the impact of institutions. *International Journal of Social Economics*, *25*(6/7/8), 1310-1325.

Saner, M., & Wilson, J. (2003). Stewardship, good governance and ethics. *Institute on Governance Policy Brief No. 19.*

Schwab, K. (2017). Insight Report: The Global Competitiveness Report 2017-2018. Retrieved from http://www3.weforum.org/docs/GCR2017-2018/05FullReport/TheGlobalCompetitivenessReport2017%E2%80%932018.pdf

Tayleur, C., Balmford, A., Buchanan, G. M., Butchart, S. H. M., Corlet Walker, C., Ducharme, H., . . . Phalan, B. (2018). Where are commodity crops certified, and what does it mean for conservation and poverty alleviation? *Biological Conservation, 217*, 36-46. doi:10.1016/j.biocon.2017.09.024

The Governance Institute of Australia (n.d.). Good Governance Guide – Key Performance Indicators (KPIs): both financial and non-financial. Retrieved from https://www.governanceinstitute.com.au/

Thorlakson, T., Hainmueller, J., & Lambin, E. F. (2018). Improving environmental practices in agricultural supply chains: The role of company-led standards. *Global Environmental Change, 48*, 32-42. doi:10.1016/j.gloenvcha.2017.10.006

Transparency International (n.d.). Corruption Perceptions Index 2017: Technical Methodology Note. Retrieved from https://www.transparency.org

United Nations Economic and Social Commission for Asia and the Pacific (n.d.). What is Good Governance? Retrieved from http://www.unescap.org/sites/default/files/good-governance.pdf

United Nations Forum on Sustainability Standards (2013). Voluntary Sustainability Standards. Today's Landscape of Issues & Initiatives to Achieve Public Policy Objectives. New York: UNFFS. Retrieved from https://unfss.files.wordpress.com/2013/02/unfss_vss-flagshipreportpart1-issues-draft1.pdf

United Nations Global Compact (n.d.). Information on the United Nations Global Compact. Retrieved from https://www.unglobalcompact.org/

United Nations Global Compact (n.d.). Information on the Global Compact Self Assessment Tool. Retrieved from http://globalcompactselfassessment.org/

UTZ (2018). Information on UTZ Certified. Retrieved from https://utz.org/

Vanhanen, T. (2000). A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research*, *37*(2), 251-265.

Van de Walle, S. (2006). The state of the world's bureaucracies. *Journal of Comparative Policy Analysis*, *8*(4), 437-448.

Van Wyk, L. V., & Chege, L. U. C. Y. (2004). Globalisation, corporate governance and the construction industry.

Weiss, T. G. (2000). Governance, good governance and global governance: conceptual and actual challenges. *Third world quarterly*, *21*(5), 795-814.

Woods, N. (2000). The challenge of good governance for the IMF and the World Bank themselves. *World development*, *28*(5), 823-841.

Zattoni, A., & Cuomo, F. (2008). Why adopt codes of good governance? A comparison of institutional and efficiency perspectives. *Corporate governance: an international review*, *16*(1), 1-15.

Appendix

Appendix 1 – Blank version of the assessment framework and comments on future changes/improvements ('Sources' column)

Categories, Sub-Categories and Criteria				
Legitimacy and Voice	Indicators	Availability	Score	Sources
Appropriate basis				
1. Share of academics among the employees in the standard setting organisation (if subsidiaries exist, only the international level)	1 = 20%+ of employees with an academic degree 0.5 = 10%+ 0 = less than 10%	Information available pre-communication/post-communication/is disclosed or does not exist		It may be difficult to predict how appropriate these estimated per centages are. This criterion could not be scored in any of the first-round assessments.
2. Accreditation authority	1 = an independent third-party organisation has the authority to conduct accreditation 0.5 = the (a department/partner of the) organisation conducts accreditation on its own 0 = no, there is no accreditation	Information available pre-communication/post-communication/is disclosed or does not exist		
3. Auditing authority	1 = an independent third-party organisation has the authority to conduct auditing 0.5 = (a department of/partner of the) the organisation conducts auditing 0 = there is no auditing	Information available pre-communication/post-communication/is disclosed or does not exist		
General quality				
4. Research activity	1 = permanent research group/projects within the organisation 0.75 = temporary research groups/projects within the organisation 0.5= independent permanent research groups/projects for the organisation 0.25= independent temporary research groups/projects for the organisation 0 = no research initiated by the organisation	Information available pre-communication/post-communication/is disclosed or does not exist		It is debateable whether independent but permanent activities are worse or better than temporary activities of the organisation.
5. Evaluation of ethics	1 = organisation states that it evaluates its ethical perspective and provides evidence (e.g. reports) 0.5 = the organisation states that it evaluates its ethical perspectives but does not provide evidence 0 = there is no evaluation	Information available pre-communication/post-communication/is disclosed or does not exist		The criterion might not be applicable as none of the standards organisations seem to think this is a necessity The 0.5 score is a bad indicator score as it overlaps with criterion 24 This criterion could not be scored in any of the first round assessments. Alternatively could e.g. check for cases where ethical rules are not respected

6. Recognition of the issue focused on by the standard	1 = more than 3 relevant actors (actors with relevance on the international level e.g. UN, EU, governments, government institutions, international trade organisations, international NGOs) provide a statement, recognising the issue as a problem that requires action0.5 = 2 or fewer actors provide a statement, recognising the issue as a problem that requires action0 = the issue is not recognised by other actors	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion appears to be fulfilled rather easily and could be evaluated with a little more depth.
7. Relative added value of the standard	1 = the standard claims to add value by explaining differences towards other standards of the same sector/it is the only standard in the sector 0 = the standard does not claim to add value or explains this	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion is problematic if the organisation, as was the case for the UTZ/RA, shies away from setting itself apart even when it clearly has more benefits. Expanding this criterion to allow a check for additional information might help.
Answerability - Formal involvement				
8. Procedure to engage stakeholders	1 = procedure exists and evidence of its application in form of e.g. reports is available 0.5 = procedure exists but no evidence is published 0 = no statement on the existence of any procedure for stakeholder engagement	Information available pre-communication/post-communication/is disclosed or does not exist		The 0.5 score is a bad indicator score as it overlaps with criterion 24
9. Coverage of stakeholder groups (geographical)	1 = minimum 40% from the global north/south 0.5 = minimum 20% from the global north/south 0 = less than 20% from the global north/south	Information available pre-communication/post-communication/is disclosed or does not exist		
10. Coverage of stakeholder groups (categories)	1 = more than 80% of the actor groups from the chain of custody are represented in the organisation 0.5 = more than 50% 0 = less than 50%	Information available pre-communication/post-communication/is disclosed or does not exist		
Direction	Indicators	Availability	Score	Sources
Goals and targets				
11. Precision of the vision	1 = the vision provides specific quantitative targets and/or milestones 0.5 = the vision provides neither milestones nor quantitative targets 0 = there is no statement on their vision	Information available pre-communication/post-communication/is disclosed or does not exist		
12. Time frame of the strategy	1 = the strategy covers the next 10+ years 0.5 = the strategy covers less than the next 10 years 0 = there is no statement on their strategy	Information available pre-communication/post-communication/is disclosed or does not exist		It may be difficult to determine how appropriate these estimated numbers are in reality. None of the organisations has a long-term oriented strategy with a lot of detail.
13. Self-assessment	1 = reoccurring self-assessment through a procedure checking the same/similar features (e.g. key performance indicators) 0.5 = self-assessment exists in some other form 0 = no statement on the existence of a self-assessment	Information available pre-communication/post-communication/is disclosed or does not exist		

Theory of change				
14. Theory of change	1 = all five theory of change building blocks are articulated; vision of success, pre-conditions, strategies and interventions, impact pathways, underlying assumptions 0.5 = at least 3 of the five building blocks are clearly articulated 0 = less than 3 of the five building blocks are clearly articulated	Information available pre-communication/post-communication/is disclosed or does not exist		
15. Empirical validation of the theory of change aspects	1 = the assumptions and causal relationships of the theory of change are based on scientific research and evidence (references) 0 = the assumptions and causal relationships are not based on scientific research and evidence / no research and evidence is referenced	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion could not be scored in any of the first round assessments. As in the case of the FSC, the strategy might be based on evidence whilst the theory of change, although based on the strategy, refers to none of that.
Performance	**Indicators**	**Availability**	**Score**	**Sources**
Progress towards goals and targets				
16. Bureaucratic delays; delay between application and successful certification (if auditing demands no changes)	1 = 30 or fewer working days 0.5 = 30-60 working days 0 = 60+ working days	Information available pre-communication/post-communication/is disclosed or does not exist		There is a risk of differences based on how this process works in the organisations. It may be difficult to determine how appropriate these estimated numbers are in reality. This criterion could not be scored in any of the first round assessments.
17. Support programmes for producers	1 = permanent support programmes/projects for producers 0.5 = temporary support programmes/projects for producers 0 = no support programmes/projects for producers	Information available pre-communication/post-communication/is disclosed or does not exist		Group certification is supportive for producers but not picked up at any point of this assessment. This might be the right place to involve this aspect in the future.
18. Relative impact of the standard	1 = total quantity of certified products is greater than that of any other standard in the same sector 0.5 = total quantity of certified products is at least 50% of the biggest standard in the same sector 0 = total quantity of certified products is less than 50% of the biggest standard in the same sector	Information available pre-communication/post-communication/is disclosed or does not exist		
19. Usage of the self-assessment (see criterion 13)	1 = reporting the results of the self-assessment and actively work on critical aspects identified by the self-assessment 0.5 = reporting the results of the self-assessment 0 = they do not report on the self-assessment	Information available pre-communication/post-communication/is disclosed or does not exist		
Responsiveness - Actual involvement				
20. Consistent formal dialogue with stakeholders	1 = evidence for consistent dialogue without the exclusion of any stakeholders 0 = no claim on the existence of consistent dialogue	Information available pre-communication/post-communication/is disclosed or does not exist		
21. Stakeholder attendance - Attendance of listed stakeholders to meetings aiming at change (decision-making, consultation for the purpose of decision-making)	1 = 70%+ total attendance for the last 5 meetings or meetings in the last 5 years (or less, depending on data availability) 0.5 = more than 35% 0 = less than 35%	Information available pre-communication/post-communication/is disclosed or does not exist		

Accountability and Transparency	Indicators	Availability	Score	Sources
Transparency - Availability of information				
22. Transparency of auditor activities	1 = standard setting organisations/accreditor receive information on day-to-day activities of the auditor (e.g. publish information on what producers receive certification) 0 = no information is shared	Information available pre-communication/post-communication/is disclosed or does not exist		
23. Standard criteria, indicators, rating system etc.	1 = the method of the standard is publicly available (indicators, weighting system etc.) 0.5 = the method of the standard is available on request (indicators, weighting system etc.) 0 = the method is not available	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion appears to be fulfilled rather easily and could be evaluated with a little more depth.
24. Public availability of information (see column D)	1 = 70%+ of the information is publicly available 0.75 = 70%+ of the information is available (publicly +on request) 0.5 = 50%+ of the information is publicly available 0.25 = 50%+ of the information is available (publicly +on request) 0 = less than 50% of the information is available	Information available pre-communication/post-communication/is disclosed or does not exist		It may be difficult to determine how appropriate these estimated numbers are in reality. More information than expected was available (between 72-92% of the criteria could be scored after the first assessment round)
Internal accountability				
25. Division of executive power	1 = executive power lies with 3 or more persons 0.5 = executive power lies with 2 persons 0 = executive power lies with 1 person	Information available pre-communication/post-communication/is disclosed or does not exist		It may be difficult to determine how appropriate these estimated numbers are in reality. Most standard setting organisations have a board with numerous members that holds the executive power.
26. Election of board members	1 = board member terms are shorter than 3 years 0.5 = board member terms are longer than 3 years 0 = board members are elected indefinitely	Information available pre-communication/post-communication/is disclosed or does not exist		
27. Accountability of the accreditor	1 = organisation has a control/review process and checks each accreditor at least once every 2 years 0.5 = organisation has a control/review process and checks each accreditor every 2 years or more 0 = organisation does not have a control/review process	Information available pre-communication/post-communication/is disclosed or does not exist		
28. Accountability of the auditor	1 = organisation has a control/review process and checks each auditor at least once every 2 years 0.5 = organisation has a control/review process and checks each auditor every 2 years or more 0 = organisation does not have a control/review process	Information available pre-communication/post-communication/is disclosed or does not exist		

External accountability				
29. Reporting mechanism	1 = existence of a formal reporting mechanism for incidents and disagreements for stakeholders 0 = no statement on the existence a reporting mechanism	Information available pre-communication/post-communication/is disclosed or does not exist		
30. Exchangeability of stakeholder representatives	1 = stakeholders decide without limitations on the individual to interact with the standard setter 0.5 = standard setter may create prohibiting conditions for chosen individuals who interact on behalf of the stakeholder 0 = standard setter may prohibit chosen individuals to interact on behalf of the stakeholder	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion appears to be fulfilled rather easily and could be evaluated with a little more depth.
Fairness	**Indicators**	**Availability**	**Score**	**Sources**
Equity				
31. Equity policies			#DIV/0!	
31.1 Employee protection/whistle-blower policy	1 = policy publicly explained (e.g. on website) 0.5 = policy explained but information only available on request 0 = policy is not explained	Information available pre-communication/post-communication/is disclosed or does not exist		The 0.5 score is a bad indicator score as it overlaps with criterion 24 This criterion could not be scored in any of the first round assessments.
31.2 Employment policy	1 = policy publicly explained (e.g. on website) 0.5 = policy explained but information only available on request 0 = policy is not explained	Information available pre-communication/post-communication/is disclosed or does not exist		The 0.5 score is a bad indicator score as it overlaps with criterion 24 This criterion could not be scored in any of the first round assessments.
31.3 Equality policy	1 = policy publicly explained (e.g. on website) 0.5 = policy explained but information only available on request 0 = policy is not explained	Information available pre-communication/post-communication/is disclosed or does not exist		The 0.5 score is a bad indicator score as it overlaps with criterion 24
32. Equality			#DIV/0!	
32.1 Female/male director balance	1 = minimum 40% male/female directors 0.5 = minimum 30% male/female directors 0 = less than 30% male/female directors	Information available pre-communication/post-communication/is disclosed or does not exist		The gender balance may not directly reflect inequality + there are other forms of inequality that can simply not be addressed by limiting this sub-category to two criteria
32.2 Female/male workforce balance	1 = minimum 40% male/female employees 0.5 = minimum 30% male/female employees 0 = less than 30% male/female employees	Information available pre-communication/post-communication/is disclosed or does not exist		The gender balance may not directly reflect inequality + there are other forms of inequality that can simply not be addressed by limiting this sub-category to two criteria
33. Employee communication			0	
33.1 Dialogue between workers and management	1 = official meetings for workers (e.g. through a committee) to communicate with the management at least once every 2 months 0.5 = official meetings for workers (e.g. through a committee) communicate with the management less than once every 2 months 0 = no official meetings for workers to communicate with the management	Information available pre-communication/post-communication/is disclosed or does not exist		May not be a relevant criterion if informal dialogue dominates and conflict is rare

34. Employee retainment			#DIV/0!	
34.1 Employee turnover rate	1 = employee turnover rate below 10% 0.5 = employee turnover rate between 10-20% 0 = employee turnover rate above 20%	Information available pre-communication/post-communication/is disclosed or does not exist		A turnover rate of 10% or less can be either desirable as it could stand for stability or it is not desirable as it may indicate stagnancy
34.2 Training and development opportunities	1 = the organisation provides employees with 2 or more development opportunities per year (examining the last 2 years, if information is available) 0.5 = ...1 development opportunity per year 0 = ...no development opportunities	Information available pre-communication/post-communication/is disclosed or does not exist		
Rule of law				
35. Accreditation: threats to the rule of law	1 = adherence to the rule of law is an accreditation criterion (e.g. checking for corruption risks) 0 = adherence to the rule of law is not an accreditation criterion	Information available pre-communication/post-communication/is disclosed or does not exist		
36. Auditing: threats to the rule of law	1 = adherence to the rule of law is an auditing criterion (e.g. checking conformity with national regulation) 0 = adherence to the rule of law is not an auditing criterion	Information available pre-communication/post-communication/is disclosed or does not exist		
37. Support for a fair and just environment	1 = organisation is part of a project/collective that explicitly aims at establishing/maintaining a just environment 0 = organisation is not part of a project/collective explicitly aims at establishing/maintaining a just environment	Information available pre-communication/post-communication/is disclosed or does not exist		This criterion could not be scored in any of the first-round assessments.
38. Internal mechanisms to adhere to the rule of law	1 = adherence to the rule of law guaranteed through policies, procedures, special departments or other mechanisms 0 = no additional mechanism to guarantee adherence to the rule of law in place	Information available pre-communication/post-communication/is disclosed or does not exist		

Appendix 2 - User guide

	User Guide
	Legitimacy and Voice
	Appropriate basis
1	Check for the share of employees with an academic degree, meaning a bachelor's degree or equivalent/higher one. This is limited to employees from the highest/international organisation that belongs to the system of the standard and takes care of decision-making and related day-to-day activities and includes all legal entities belonging to the international body but excludes subsidiary organisations.
2	Check whether the bodies responsible for the auditing procedure are enabled to do so by an independent accreditation authority, through accreditation or another procedure granting them the authority to conduct audits. It would be sufficient for this criterion of the standard setter demands auditors to obtain, for example, ISO certification.
3	Check whether the auditing procedure is facilitated by organisations that are independent from the standard setting organisation, meaning that they may, for example, certify other standards at the same time or discontinue auditing for the standard setter.
	General quality
4	If there is no group/actor dedicated to research activities, check whether any of the projects are aiming at producing research results. It is important that the standard setting organisation initiated such research. Independent research published on activities of the standard does not count towards this criterion.
5	Check resource/data bases of the organisation for information on such an evaluation. It would suffice if a review of their own ethics is part of the self-assessment (criterion 13)
6	Check for actors who determined that the issue that is stated in the vision of the standard is problematic. 'Relevant' actors are actors who can have an impact on the international level e.g. UN, EU, governments, international trade organisations, international NGOs
7	Check if the standard (on their own!) explains its own role compared to standards in the same sector or whether it clearly makes statements on features that are exclusive to the standard being examined; e.g. "This standard is the only one considering XX", "Our standard ranks first in XX"
	Answerability - Formal involvement
8	Check whether or not the organisation explains its method (e.g. consultations, invitations to meetings, membership in a certain organ...) to interact with its stakeholders. For this criterion the depth of such interaction and who establishes contact do not matter.
9	Check whether listed stakeholders are from the global north or global south. 'Listed stakeholders' are those who are invited to important meetings of the standard setter or who are actual members of the system. If neither of these two (meetings or membership) exist, this criterion is not applicable.
10	Check whether listed stakeholders (minimum of 1) are from all parts of the chain of custody (generally: ground level producer, intermediary buyer, retailer, consumer). 'Listed stakeholders' are those who are invited to important meetings of the standard setter or who are actual members of the system. If there is no interaction like this the criterion is not applicable.
	Direction
	Goals and targets
11	Check the vision itself but also any relevant documents that may describe the strategy (typically a single e.g. "strategic plan") to reach the goals
12	Check the vision itself but also any relevant documents that may describe the time frame of the vision/strategy
13	Check for the existence of a self-assessment method that clearly aims at improving certain aspects of the standard's system or just the standard itself. To achieve the full score, this assessment should check the same or very similar features every time it is undertaken. This means that a re-occurring review based on current issues is insufficient.
	Theory of change
14	Similar to criteria 11 and 12, the strategic vision document (if it exists) can provide insight into these aspects. Some organisations also explicitly describe their theory of change (e.g. ISEAL requires its members to have a theory of change)
15	Check whether scientific sources are used (e.g. cited for statements) for important assumptions of the vision and causal relationships.
	Performance
	Progress towards goals and targets
16	It may be necessary to contact the auditor for this information. Depending on the availability of information, ideally the average of the five most recent cases is taken. If this is not possible any number of recent cases would have to suffice but should be noted.
17	Any kind of project, programme or activity can suffice to judge this criterion, as long as it is aimed at going beyond the ability of producers to just fulfil the standard's requirements. Permanence is established if some projects are established indefinitely or if there is a guarantee (e.g. through a fund or a certain department) for a consistent stream of projects.
18	Comparison of the volume (alternatively 'area' if there is a lack of data) of certified material/products. If a standard certifies multiple products, this score is created for each of those products and then averaged.
19	It must be clear how the self-assessment contributes to some kind of change or how it will be used in the future. Finding actual proof of a future application of self-assessment results is not necessary; a claim may suffice to judge this criterion. If there is no self-assessment method (criterion 13) this criterion is not applicable.
	Responsiveness - Actual involvement

20	As long as dialogue occurs regularly and in the same manner this criterion may be judged as sufficient. Consistency is defined by the reoccurrence of the same form of dialogue as opposed to dialogue that is only initiated in certain situations (e.g. complaints). While time intervals are not considered by this criterion, the dialogue should ideally occur more than every few years. Opportunities to meet up with the standard setter (see criterion 21) should not count towards this criterion as it relies on the initiative of the stakeholder to get involved on their own rather than the standard setter should more actively try to communicate with its stakeholders.
21	If there are various opportunities for stakeholders to meet up with the standard setter, the most impactful and universal one should be examined. Preferably it should be a meeting where important decisions are made and all stakeholders are invited. 'Listed stakeholders' are those who are invited to important meetings of the standard setter or who are actual members of the system. If there are no such stakeholders (criteria 9 and 10) then this criterion is not applicable.
	Accountability and Transparency
	Transparency - Availability of information
22	This criterion examines the communication between auditors and standard setter/accreditor. While more constant dialogue would be better, this criterion cab be judged as sufficient if the auditor provides a report for every applicant.
23	Check availability of details on the criteria, indicators etc. of the standard.
24	Any criteria filled through the examination of public sources or pointed at (but still publicly available) in the communication will be considered "publicly available" Any criteria filled through additional information provided by the organisation by communicating with them will be counted towards "available on request"
	Internal accountability
25	Quantity of people involved in the day-to-day decision-making and organisational as well as standard changes. This typically is the quantity of board members but depends on the power than is granted to the board.
26	Check the duration of the board members terms. Potential re-election does not count towards this duration.
27	Checking whether the standard setter checks the accreditor's work through a specific procedure or other mechanism. This criterion is satisfied if the accreditor has to regularly provide information on its interaction with the auditors that it grants accreditation.
28	Checking whether the standard setter checks the auditor through a specific procedure or other mechanism. This criterion is satisfied if the auditor has to go through a certain procedure that involves a check-up with active involvement of the accreditor. Supplying reports does not satisfy this criterion.
	External accountability
29	Check for a distinct method to report incidents, complaints et cetera that is available to anyone and clearly separated from the typical contact channels. While a detailed description of the procedure would be insightful, it suffices to just find and reference this mechanism.
30	If stakeholders become listed stakeholders (see criterion 9/10 for a description) and register certain individuals as representatives, check whether or not there are any restrictions on who these representatives may be.
	Fairness
	Equity
31.1	Check resource/data bases of the organisation for information on the policies.
31.2	Check resource/data bases of the organisation for information on the policies. If elaborate information on the employment procedure exists (e.g. in job listings), it may substitute legal documents focusing on the employment policy.
31.3	Check resource/data bases of the organisation for information on the policies. If elaborate information on the stance towards equality (also for others within their system e.g. farmers or ground level producers) exists, may substitute legal documents focusing on equality.
32.1	Male and female board members from the highest/international organisation that belongs to the system of the standard and takes care of decision-making and related day-to-day activities.
32.2	Male and female employees from the highest/international organisation that belongs to the system of the standard and takes care of decision-making and related day-to-day activities. This includes all legal entities belonging to the international body but excludes subsidiary organisations. If information is limited this score may be based on the staff/team that is presented on the website or elsewhere.
33.1	Check resource/data bases of the organisation for information on a formal dialogue between employees with its management. This criterion is limited to the context of the (international) standard setting organisation and not related to others in the system such as auditors, subsidiaries etc.
34.1	Check for the quantity of individuals who got employed and those who left employment at the (international) standard setting organisation and all of its legal entities. Calculation of incoming employees - outgoing employees, based on data from the most recent year.
34.2	Check for training and development opportunities provided by the (international) standard setting organisation and its legal entities for any employees in the system (incl. auditors), excluding those already covered by criterion 17 (producers).
	Rule of law
35	Check the accreditation requirements for auditors for an explicit demand for conformity of the auditor with the law. It is sufficient if this criterion exists, without an extensive procedure to check the legality of their activities.
36	Check the standard's criteria for an explicit demand for conformity of the producer with the law. It is sufficient if this criterion exists, without an extensive procedure to check the legality of their activities.
37	Support for a fair and just environment is established if the organisation is part of a collective that explicitly aims at the establishment/maintenance of a just environment or if it promotes projects that explicitly aim at eradicating illegal/unjust activities. This criterion refers to legal aspects and should not check for aspects like gender equality.
38	Check resource/data bases of the organisation for information on the policies. If there is a department/policy/other mechanism that emphasizes/checks the adherence to the law of the standard setting organisation's activities itself, this criterion is satisfied.

Appendix 3 – Information Boxes on the used initiatives

The Database of Political Institutions
Total of 108 variables for
1. Chief executive variables
2. Party variables in the legislature
3. Electoral rules
4. Stability and checks & balances
5. Federalism

Box 1 – Categories of The Database of Political Institutions

Research by Henisz (2010) - Political Constraint Dataset (POLCON)
Based on country risk indices
1. Number of independent veto points over policy outcomes
2. Distribution of preferences of the actors that inhabit them
combined with various indicators for growth

Box 2 – Indicators of the Political Constraint Dataset by Henisz (2000)

Polity IV
Institutionalised democracy and institutionalised autocracy leading to the combined polity score and revised combined polity score
Component variables:
1. Competitiveness of executive recruitment (XRCOMP)
2. Openness of executive recruitment (XROPEN)
3. Constraints on chief executive (XCONST)
4. Regulation of participation (PARREG)
5. Competitiveness of participation (PARCOMP)

Box 3 – Component variables of Polity IV

Research by Vanhanen (2000) - Polyarchy
1. Participation measured by % of population who vote
2. Competition measured by % share of total votes among small political parties

Box 4 – Indicators of Polyarchy by Vanhanen (2000)

Worldwide Governance Indicators
1. Voice and Accountability
1.1 Economist Intelligence Unit (Initiative)
1.1.1 Orderly transfers
1.1.2 Vested interests
...
1.2 Freedom House (Initiative)
1.2.1 Civil liberties
1.2.2 Political rights
...
2. Political Stability & Absence of Violence/Terrorism
3. Government Effectiveness
4. Regulatory Quality
5. Rule of Law
6. Control of Corruption

Box 5 – Categories and (some) sub-categories of the Worldwide Governance Indicators

World Economic Forum's Global Competitiveness Report
1st pillar: Institutions
Public institutions
1. Property rights
2. Ethics and corruption
3. Undue influence
4. Public-sector performance
5. Security
Private institutions
6. Corporate ethics
7. Accountability
2nd pillar: Infrastructure
3rd pillar: Macroeconomic environment
4th pillar: Heath and primary education
5th pillar: Higher education and training
6th pillar: Goods market efficiency
7th pillar: Labour market efficiency
8th pillar: Financial market development
9th pillar: Technological readiness
10th pillar: Market size
11th pillar: Business sophistication
12th pillar: R&D innovation

Box 6 – Pillars and (some) categories/criteria of the World Economic Forum's Global Competitiveness Report

Millennium Challenge Account
Ruling Justly
1. WGI Control of Corruption (WGI CC)
2. WGI Government Effectiveness (WGI GE)
3. WGI Rule of Law (WGI RL)
4. WGI Voice and Accountability (WGI VA)
5. Political Rights (FH PRI)
6. Civil Liberties (FH CLI)
Investing in People
7. Immunisation Rates
8. Public Expenditure on Primary Education
9. Public Health Expenditure
10. Primary Girls' Education Completion Rate
11. Natural Resource Management
Economic Freedom
12. Inflation Rate
13. Fiscal Policy
14. Trade Policy
15. WGI Regulatory Quality (WGI RQ)
16. Business Start-Up
17. Land Rights and Access

Box 7 – Categories and sub-categories of the Millennium Challenge Account

Corruption Perceptions Index
1. African Development Bank Country Policy and Institutional Assessment 2016
2. Bertelsmann Stiftung Sustainable Governance Indicators 2017
3. Bertelsmann Stiftung Transformation Index 2017-2018
4. Economist Intelligence Unit Country Risk Service 2017
5. Freedom House Nations in Transit 2017
6. Global Insight Country Risk Ratings 2016
7. IMD World Competitiveness Center World Competitiveness Yearbook Executive Opinion Survey 2017
8. Political and Economic Risk Consultancy Asian Intelligence 2017
9. The PRS Group International Country Risk Guide 2017
10. World Bank Country Policy and Institutional Assessment 2017
11. World Economic Forum Executive Opinion Survey 2017
12. World Justice Project Rule of Law Index Expert Survey 2017-2018
13. Varieties of Democracy (V-Dem) 2017

Box 8 – Initiatives used by the Corruption Perceptions Index

IoD Good Governance Index
Stakeholder Relations
1. CSR/sustainability committee
2. Employee protection / whistle-blower policy?
3. Is the company a signatory to the Prompt Payment Code?
4. Most Admired Companies
5. UN Global Compact signatory
6. GRI criteria compliance
7. Does the company explain how it engages with its stakeholders?
8. Does the company claim to have an ISO 9000 certification?
Board Effectiveness
1. Separate CEO and chairman
2. Independent chairman
3. % of non-executive directors on the board
4. % of female directors on the board
5. % of non-British directors on the board
6. Fewer than eight or more than 15 directors
7. Board of directors age range
8. Number of board meetings held
9. % of board meeting attendance
10. % of directors on the board longer than nine years
Audits and Risk/External Accountability
1. % of independent directors on audit committee
2. % of audit committee meeting attendance
3. Years with current audit company
4. Ratio of fees for non-audit/audit work to auditors
5. Auditor's report is qualified/unqualified?
6. Is the company in the process of a material earnings restatement?
7. RepRisk (provided via CQI)
8. Size of audit committee
9. Number of audit committee meetings

Box 9 – Indicators and categories of the Institute of Director's Good Governance Index

UN Global Compact Self Assessment Tool
Human rights
1. Principle 1 - Businesses should support and respect the protection of internationally proclaimed human rights
2. Principle 2 - Businesses should make sure that they are not complicit in human rights abuse
Labour
3. Principle 3 - Businesses should uphold the freedom of association and the effective recognition of the right to collective bargaining
...
Anti-corruption
4. Principle 10 - Businesses should work against corruption in all its forms, including extortion and bribery

Box 10 – Relevant categories and principles of the UN Global Compact Self Assessment Tool

Robecosam Survey
1. Social dimension
2. Economic dimension
3. Customer relationship management
4. Anti-crime policy

Box 11 – Categories of the Robecosam Survey

EO100 Standard
Principle 1: Corporate Governance, Accountability & Ethics
1.1 Legal Compliance
1.2 Governance, Policies & Codes Of Conduct
1.3 Fiduciary Compliance & Disclosure
1.4 Bribery & Corruption
1.5 Money Laundering/ Terrorism
1.6 Transparency & Disclosure
Principle 2: Human Rights, Social Impact & Community Development
Principle 3: Fair Labor & Working Conditions
Principle 4: Indigenous Peoples' Rights
Principle 5: Climate Change, Biodiversity & Environment
Principle 6: Project Life Cycle Management

Box 12 – Principles and (some) categories of the EO100 Standard