



S.A.P.I.E.N.S

Surveys and Perspectives Integrating Environment and Society

8.1 | 2015
Vol.8 / n°1

Moving to metrics: Opportunities and challenges of performance-based sustainability standards

Michael Veale and Rafael Seixas

Nicolas Renard (ed.)



Publisher
Institut Veolia Environnement

Electronic version

URL: <http://sapiens.revues.org/1713>

ISSN: 1993-3819

Electronic reference

Michael Veale and Rafael Seixas, « Moving to metrics: Opportunities and challenges of performance-based sustainability standards », *S.A.P.I.E.N.S* [Online], 8.1 | 2015, Online since 15 December 2015, connection on 30 September 2016. URL : <http://sapiens.revues.org/1713>

The text is a facsimile of the print edition.

Licence Creative Commons

Perspectives

Moving to metrics: Opportunities and challenges of performance-based sustainability standards

Michael Veale¹ and Rafael Seixas²

¹Department of Science, Technology, Engineering & Public Policy (UCL STEaPP), University College London, Boston House 2nd Floor, 36–38 Fitzroy Square, London, W1T 3EY, Email : m.veale@ucl.ac.uk

²Bonsucro, 50-52 Wharf Road, London N1 7EU, Email : rafael@bonsucro.com

Abstract

The rise of global sustainability standards has led to an energetic discussion about their consequences and outcomes. Almost all standards today are built around ‘technology-based’ indicators, which prescribe certain practices assumed to lead to sustainable outcomes. However we are now seeing the emergence of the first ‘performance-based’ metric sustainability indicators, directly measuring outcomes without prescribing particular methods to reach them. This paper presents the example of the Bonsucro Production Standard, a sustainability standard for the sugarcane sector, and identifies five relevant areas opened up by performance-based metrics. These are flexibility in application, provision of information, the creation of dynamic standards, the enabling of adaptive management, and the harmonisation of policy instruments. Opportunities and challenges within each area are discussed in relation to a wide literature from a variety of disciplines, informing opportunities for standard-systems to explore within their own activities, as well as an agenda for future research.

Keywords: sustainability, standards, performance-based standards, metric standards, indicators, agricultural sustainability standards, certification, labeling, eco-labelling, labelling, eco-labeling, environmental standards, social standard, private governance, market governance, adaptive management, sugarcane, dynamic standards, technology-based standards, sustainable agriculture, visualization, visualisation

TABLE OF CONTENTS

1. Introduction
2. Example of a performance-based standard: the Bonsucro Production Standard
3. Relevant new concepts arising from performance-based metrics
 - 3.1 Flexibility in application
 - 3.2 Information provision
 - 3.3 Dynamic standards
 - 3.4 Adaptive management
 - 3.5 Instrument harmonisation
4. An agenda for future research
5. Conclusion

1. INTRODUCTION

Environmental and social problems causing concern at a global scale have resulted in the development of a broad range of sustainability standards. Although not exclusively 'private', these have for the most part been developed by a range of private actors, especially within the NGO and business sectors. Many factors drive the uptake of certification schemes: many consumers now share stronger post-materialistic concerns (Inglehart, 1997); firms seek to build legitimacy and supply chain power (Bernstein & Cashore, 2007); upstream producers seek preferable prices and markets for their products (Jaffee, 2007; Neilson & Pritchard, 2009); NGOs seek funding, publicity and influence (Schwesinger Berlie, 2010); and a whole range of actors attempt to reduce their risk and exposure as they find themselves within uncertain socio-ecological and socio-economic systems (Beck, 1992; Giddens, 2002). Indeed, over 450 'ecolabels' across 25 industry sectors are now recorded by the *Ecolabel Index*¹, a global directory of sustainability labels.

This field has a new and developing lexicon, which benefits from clarification. This paper follows Matus (2010: 80). A *standard* lists 'specifications and/or criteria for the manufacture, use, and/or attributes of a product, process, or service'. *Certification* is the 'process, often performed by a third party, of verifying that a product, process or service adheres to a given set of standards and/or criteria'. *Labelling* is the 'method of providing information on the attributes, often unobservable, for a product, process or service'. While this paper is primarily concerned with agricultural commodity certification, the terms and discussion are more widely generalisable.

Currently, global sustainability standards largely consist of *technology-based* indicators. Technology-based standards prescribe certain technology—"knowledge of how to fulfil certain human purposes in a specifiable and reproducible way" (Brooks, 1980: 66). The standard represents a hypothesis—the technology is expected to promote certain desired outcomes. These hypotheses are often drawn from 'best practices' and guidance. Their expected consequences stem from evidence ranging from anecdotes to large-scale randomised control trials.

There is increasing interest in using *performance-based* metric standards in place of technology-based indicators within the context of sustainability. These attempt to measure outcomes directly, rather than proxying them with practices. *Bonsucro*, one of the first global performance-based metric standards, will be discussed below. Other initiatives, such as the *Sustainable Apparel Coalition* in the clothing sector, are also developing sustainability metrics, although these have yet to crystallise into standards.

What are the benefits and pitfalls of performance-based metric standards in the field of sustainability? First, we present the case of *Bonsucro* as an emerging performance-based standard. Then, we draw on diverse strands of

literature to point to five fields in which metric standards are promising: flexibility in application, provision of information, creating dynamic standards, enabling adaptive management, and harmonisation of policy instruments.

2. EXAMPLE OF A PERFORMANCE-BASED STANDARD: THE BONSUCCRO PRODUCTION STANDARD

Bonsucro, initially the Better Sugarcane Initiative, is a certifier and standard-setter for sugarcane and derivative products. Founded in 2008, it awards certificates to mills that meet its *Production Standard*, with the agricultural systems and lands that supply the raw cane falling under the same certificate. *Bonsucro* certifies a total of 4.08% of the world's surface sugarcane production.² The secretariat is based in London, given the time zones spanned by sugarcane-producing countries, with additional ground staff in Brazil. The secretariat organises the training and accreditation of recognised auditing organisations. Certification is also required for those who handle certified produce along the supply chain, called 'chain of custody' certification.

Bonsucro is primarily a performance-based certification, requiring facilities seeking certification to input 237 data points about their activities. Most of these are metric, although some of them are 'Y/N' inputs familiar from technology-based standards. Some Y/N questions are technology-based (e.g. 'Sulfitation Process Used?'), while others cover less quantifiable issues (e.g. 'The right to use the land and water can be demonstrated and is not legitimately contested by local communities with demonstrable rights'; 'Availability of sufficient drinking water to each worker present in the mill'). The breakdown of metric to non-metric data points, as well as the general structure of the certification process, is shown in Figure 1, along with (non-exhaustive) examples of some of the points in each section.

Compliance with the standard is determined using a calculator built with Microsoft Excel that is distributed to mills seeking certification. This calculator takes the provided data points, which are independently audited for veracity, and enters them into formulae to calculate 82 criteria. Certification is awarded if the 16 core criteria are met, in addition to 80% of the total individual criteria.

3. RELEVANT NEW CONCEPTS ARISING FROM PERFORMANCE-BASED METRICS

3.1 FLEXIBILITY IN APPLICATION

Technology-based standards confer a number of benefits. Many poor outcomes do indeed result from particular common and unremedied practices, and simply proposing and implementing better ones can be an easy step to a desirable socio-ecological situation. Training auditors and outreach staff to take a technology-based view—looking at and advising on practices on the ground—is often cheaper than examining outcomes, which can be methodologically complex. Many facilities seeking certification are also more comfortable with

1 ecolabelindex.com

2 Correct as of December 2015; see <http://bonsucro.com/site/in-numbers/>

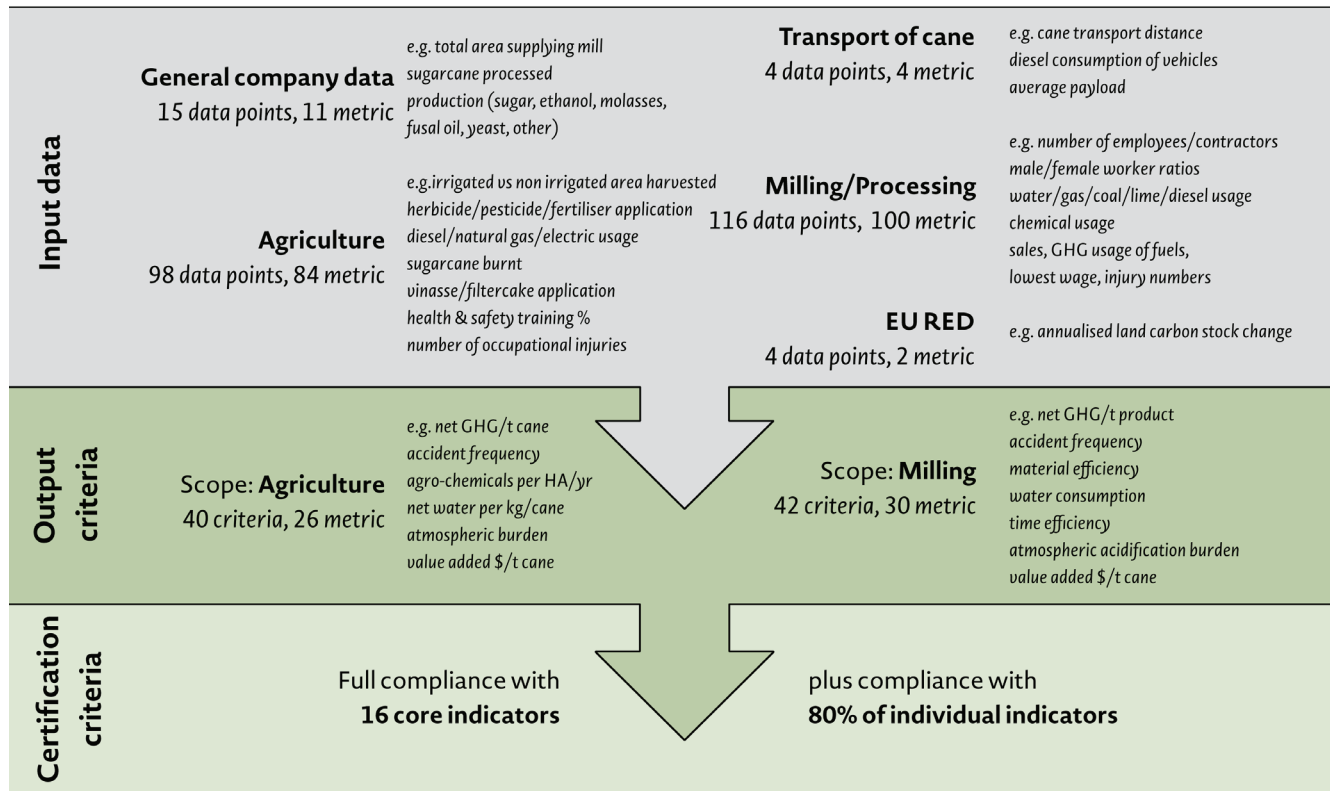


Figure 1. The structure of input data and indicators in the Bonsucro Production Standard

a technology-based view of the world, and find intangible sustainability impacts difficult to understand, explain and integrate.

However, technology-based standards also generate their own issues. Their proposed causal mechanisms, which may be more complex than anticipated, may not hold in different contexts. Top-down command-and-control measures can result in unforeseen consequences for welfare and ecosystems, as rigid measures can restrict the natural variability of socio-ecological systems that helps generate resilience against shocks (Holling & Meffe, 1996). Practices that are only tested or developed in certain contexts may have different consequences elsewhere. Limited generalisability makes the strength, reliability and even the directionality of outcomes unclear in the heterogeneous contexts usually faced by global standards. Where they are tested, they are considered in isolation rather than in the synergies we find in the real world.

Technological prescription reduces firms' flexibility to select the least burdensome method to achieve a desired aim in their particular situation (Gunningham, 1996). The imposition of given approaches can cause resentment by those unable to adapt them to their needs (Bardach & Kagan, 1982). Moreover, prescription is difficult to reconcile with the local co-creation of practices. Local co-creation is considered important for many normative and instrumental reasons: key understandings of complex socio-ecological systems are often embedded in 'local knowledge' (Berkes & Folke, 2002; Gadgil *et al.*, 1998; Gadgil *et al.*, 2003; Lansing & Kremer, 1993), natural resource management has a significant cultural dimension (Ostrom &

Nagendra, 2006), and complex sustainability problems appear to necessitate integrating society in a 'transdisciplinary' mode (Lang *et al.*, 2012).

Technology-based standards innovate as fast as the standard is updated, while performance-based standards innovate as fast as those seeking certification can keep up. Prescribed practices exhibit lags before incorporating innovation, as standards have to be renegotiated and reformed, which is both a technical and a political process. In contrast, performance-based standards offer flexibility for producers to use the best fitting technology or practice they have available (Ribaud & Caswell, 1999). It does not matter how the targets are achieved, as long as they are achieved. Yet it can also be argued that while performance-based standards provide the 'pull' to achieve outcomes, they do not provide the positive 'push'. Innovative practices are not utilized because they are suddenly understood or available—technology-based standards may produce awareness or capacity to utilize them, while performance-based standards shift awareness- and capacity-building onto the shoulders of the body being certified.

Supporting the flexibility of performance-based standards is not an easy task, and if not done carefully, could even undermine the benefits, by absolving the standard-setting body of the requirement of establishing stable and globally-relevant causal pathways between practice and outcome. Still, this new-found flexibility needs to be managed and supported. Innovation is thought to be fostered by shared cognitive spaces (Nonaka, 1994) and inter-organisational networks (Powell & Grodal, 2006). To ask facilities lacking these to innovate and

apply solutions without guidance could be problematic. In order to generate and understand locally-relevant sustainable practices, it appears facilities require networks of knowledge and expertise connecting them to other regional practitioners and experts (Carolan, 2006). On the other hand, if guidance is provided, these same facilities may fall back on them strongly (Coglianese *et al.*, 2003; Gunningham, 1996). In the absence of such knowledge production and dissemination, the guidelines requested may act similarly to binding rules, removing the benefits of performance-based regulation (Bardach & Kagan, 1982).

The required technical knowledge for the enforcement of a performance-based standard makes spanning several sectors difficult. In agriculture for example, single-commodity initiatives are able to cover all aspects of production to a much greater level of detail than multi-crop standards. It is hard to imagine a single standard that could set production performance targets for multiple crops in different climates in countries with different production realities.³ In this sense, the flexibility of a performance-based standard does not extend beyond a single good or service. In the case of expansion of performance-based standards, a key challenge will involve attempting to ensure that standards across sectors are coherent, similarly rigorous, and complementary, while recognising their necessary heterogeneity. This will entail combining numerous views, values and approaches. Given that producing standards and governance for a single sector requires a broad base of supporters and constant consultation with multiple stakeholders, recognising dispersed expertise, the necessity of local adaptation, openness to new knowledge and innovation (Abbott & Snidal, 2009), the challenge of 'orchestrating' these standards in inclusive and well-informed ways is daunting indeed.

3.2 INFORMATION PROVISION

The creation of audited metrics for a facility, capturing indicators that otherwise there might be few incentives to measure, has direct and indirect benefits beyond meeting standards. Agencies extending credit are increasingly interested in environmental degradation and corporate social responsibility. Social and environmental performance conveys important non-financial information increasingly used to evaluate creditworthiness, which can result in lower financing costs for more socially and environmentally responsible facilities (Attig *et al.*, 2013). Credible, externally audited data can serve this extra purpose, and can be a base for developing indices, such as the Dow Jones Sustainability Index, based on common metrics. Currently such indices tend to be built through the coding of sustainability reports (e.g. Morhardt *et al.*, 2002). While these have been increasingly standardised in format through adoption of guidelines such as those of the Global Reporting Initiative, harmonising the methodologies of the underlying measurements in a comparable way would allow for less subjective comparison between performance.

³ It is important not to confuse a standard with a certifier. Rainforest Alliance, for example, has different standards for different crops, even though there are commonalities where possible and convenient.

The collated information itself can be directly used as a decision support tool as well as to assess standard compliance. Significant findings, especially in the field of energy use, report that if an indicator is measured and observed by those who have influence over it, it is better managed (Faruqui *et al.*, 2010). When it comes to actually improving compliance and the propensity for desired outcomes, it appears that the best approach is a combination of enforcement—through auditing and the awarding or removing of certificates—and a 'management mechanism' that embraces a problem-solving approach based on capacity building (Tallberg, 2002).

The Bonsucro Calculator, discussed above, allows users to visualise their whole production in terms of environmental, social, and economic results, as well as being informed as to whether they comply with the standard. Through being able to understand their shortcomings and advantages in relation to the requirements of Bonsucro's metric standard, the tool may aid producers in identifying action points and prioritising policies and investments.

Standards systems, by virtue of collecting data across the sector, can also provide context to the individually collated and verified information. A wide variety of modern literature in behavioural economics and social psychology points to the roles of comparison, competition and social norms in ensuring better outcomes. Studies in fields such as energy usage (Allcott, 2011; Nolan *et al.*, 2008; Schultz *et al.*, 2007), voting (Gerber & Rogers, 2009), conservation of green spaces (Cialdini, 2003), and charitable giving (Frey & Meier, 2004) all note that providing context to individual performance or intention has a positive effect on outcomes. Informing facilities how well they perform within a distribution of their peers (which is possible with performance-based metric standards) would therefore seem likely to indirectly encourage facilities to go beyond compliance, as well as identifying areas of improvement where better practices and more positive impacts are already being carried out elsewhere, especially if they are at the lower end of the spectrum of performance (Schultz *et al.*, 2007).

The data collected in the process of certifying facilities to a performance-based standard can also be used in different ways by other interested bodies. It can be anonymously provided to researchers to investigate relevant questions and relationships of interest. It can also be scaled up to a national or regional level to answer questions about relative aggregate performance across space or time.

The increased range of use of data is an exercise in knowledge production. This involves creating products with the data that lie on the boundary between knowledge and action. Creating meaningful products that are acted upon, be they aggregated sustainability reports or localised sustainability decision support, requires standard systems to grapple with questions of how to imbue them with sufficient *salience*, *credibility* and *legitimacy* (Cash *et al.*, 2003; Mitchell *et al.*, 2006). *Salience* refers to the relevance of the product to decision-makers



and stakeholders; credibility refers to the scientific adequacy of the technical evidence and arguments; and legitimacy refers to the perception that the production of information and techniques has been respectful of divergent values and understandings, and is unbiased and fair in its conduct and treatment of opposing views. Given that an increase in one of these factors can often have a negative effect on another, drawing the balance can be difficult, and requires time, expertise, and serious effort spent on 'boundary work' (Clark *et al.*, 2011). Performance-based standards bring many new types of knowledge products, but in doing so, bring the need for credibility, salience and legitimacy to the fore.

3.3 DYNAMIC STANDARDS

Sustainability standards have to balance the requirements of the certification on one hand and the level of adoption by firms and impacts on sustainability on the other (Cashore *et al.*, 2007; Lebel, 2012). A maximum positive aggregate impact on sustainability within this rigour-versus-uptake trade-off is not easily established nor understood. Metric performance-based certification allows standards-systems to try to approach this trade-off from a few new directions.

Firstly, standards can create dynamic indicator compliance criteria. This means that rather than a set standard, formulae can be applied to make the threshold for indicator compliance contingent on characteristics of the facility. Furthermore, it also creates opportunities to link facility performance to the surrounding environment. Facility emissions, social efforts and the like are not outcomes in themselves—they too are proxies for actual impacts. Embedding performance in this way creates measurements of what Veleva *et al.* (2001) call indicators of *sustainable systems*. While this represents a steep methodological challenge, it also creates new possibilities for assessment, intervention and change.

Secondly, rules more relevant to the topic of the standard can be applied in order to make a decision on certification. Currently, standards generally certify facilities that meet core criteria, plus a given proportion of other criteria in the standard, which may or may not be weighted. However, this is only one way of many in which multi-criteria decisions can be undertaken. A whole array of potential aggregation functions for multi-criteria decision-making exist (for an overview, see Campanella & Ribeiro, 2011). For example, it is possible to build an indicator that averages values below a certain threshold differently from values above. Developments of more nuanced decision rules for certification can incorporate findings about the multiple and interlinked underpinnings behind social development (Nussbaum, 2001; Sen, 1999) and poverty (Barrett *et al.*, 2011; Carter & Barrett, 2006), as well as interlinked (Gunderson & Holling, 2002) and networked (Janssen *et al.*, 2006) environmental systems, in addition to the links becoming increasingly apparent between social and economic performance (Ambec & Lanoie, 2008; Porter & Kramer, 2011).

However, dynamic standards and decision rules also come with some key caveats. While such innovations are arguably better

at promoting sustainability, and being relevant to the socio-ecological and socio-economic systems that the standard attempts to govern, they are difficult to understand intuitively. In particular, it is difficult for consumers or facilities seeking certification to judge the stringency of such an indicator or standard if the formulae are not easy to calculate mentally. It may be possible that standards like this confuse those seeking to meet them, as it is potentially more difficult to aim at a 'moving target'. It will be up to the standard system to not only explain but also 'sell' these techniques in a way that does not undermine the perceived credibility of the standard.

Furthermore, the value-laden nature of decision-rules underpinning performance-based standards can make them difficult both to calculate and renegotiate in a way considered legitimate by stakeholders. Quantification of the social dimension of sustainability, for example, is both methodologically underdeveloped, and considered by many to be a political stance in and of itself. Reaching agreement on this front is challenging during standard-setting, and renegotiating to create dynamic standards is likely to present political challenges. Whether such a metric is possible to calculate reliably and audit repeatedly is a considerable challenge. Language and gender, for example, have been pointed to as important issues, which can prevent auditors from understanding the social realities of facilities they are inspecting (van der Wal, 2011). Issues that remain in consistent measurement cast doubt over the immediate possibility of dynamic standards in some areas.

3.4 ADAPTIVE MANAGEMENT

Adaptive management entails treating all policy decisions as hypotheses, using them to test outcomes and assumptions, and revising strategies and underlying beliefs (Gunderson *et al.*, 1995). Metric performance-based standards allow for many new opportunities to learn from data in order to ensure the integrity of certified facilities and the rigour of the standard itself.

Through looking at certified data, standard statistical methods for outlier testing can identify anomalous data, which could flag concerns about facilities or audits. Through grouping methods such as principal component analysis, or classification methods such as machine-learning algorithms, constellations of low or high performance can be discerned. This can help to identify persistent areas of strength or weakness with regards to the sustainability of certified facilities, which can lead to the creation of relevant extension or outreach work to address struggles or learn from high-flyers. On an aggregated scale, spatial and temporal performance can be examined in order to better understand the dynamics of performance across nations and regions, or throughout time.

Data can also contribute to standard revision. Revision of a standard is advisable as a best practice for many reasons⁴. All sustainability standards codify a certain view of sustainability,

⁴ See the ISEAL Code of Good Practice for Setting Social and Environmental Standards: <http://www.isealalliance.org/our-work/defining-credibility/codes-of-good-practice/standard-setting-code>

a concept that by definition is fluid and changing (Robinson, 2004). As societies' views on both the scientific and social underpinnings of the idea develop, so should sustainability standards. In addition, the subject of certification is also changing—some new social, environmental or economic problems may emerge, and some old ones may become less relevant. Given the changing and varied nature of socio-ecological systems, it is unlikely that a static approach will work for more than a brief period of time in one specific context (Blann *et al.*, 2003). A standard must also position itself in the marketplace at some position on the trade-off between rigour and uptake (Cashore *et al.*, 2004; Lebel, 2012), ideally to aim for a maximum aggregate impact on sustainability.

Data such as this can be used to identify how rigorous a standard is. Some indicators may turn out to be too difficult to meet for facilities, while others may turn out to be too easy—especially if data on actual practice was scarce at the initial standard-setting. Facilities with certain characteristics, such as those from a certain country, may easily meet indicators that facilities with different characteristics struggle with. This might indicate the need for a heterogeneous standard in order to ensure improvement across the board, especially when the relevant problems of sustainability differ across the world. Some indicators may be easily met by all facilities, perhaps indicating that they can be dropped and potentially replaced by another relevant issue where measurement is desired. None of this is prescriptive, but alongside stakeholder discussions on standard revision, it provides an extra source of information to focus deliberation and revision on where it seems warranted.

However, this can be difficult to implement for a variety of reasons. Incumbent managers appear to perceive adaptive systems as a threat to current management practice (Walters, 1997). In addition, evaluative learning that accepts failure, confronting, questioning and challenging the assumptions that preceded it, requires especially strong leadership throughout the process (Argyris, 1976). To not only learn but to act from data requires a well-designed organisation in addition to a well-designed standard.

3.5 INSTRUMENT HARMONISATION

Several sustainability certification schemes have synergised or otherwise interacted with public policy. *Leadership in Energy and Environmental Design* (LEED), a green building certification, has become incorporated into many public procurement and building codes, thus superseding national regulation (SCSKASC, 2012). *Forest Stewardship Council* certification is required by the Guatemalan Government for forestry companies that operate in the Mayan Biosphere reserve (UNCTAD, 2011). Programmes such as the *Clean Development Mechanism Gold Standard*, or *Bonsucro EU Standard*, 'raise the bar' on environmental criteria for projects or commodities while also meeting public policy requirements such as the CDM credit programme or the EU Renewable Energy Directive respectively (Fortin & Richardson, 2013; SCSKASC, 2012). Collecting data for standards can also allow that data to be available to comply with policy at other

levels and for other purposes, thus economising on the costs of collection and auditing.

Performance-based standards can both 'upload' and 'download' metrics and methodologies to and from public policy. Commonly used methodologies can be 'downloaded' from those developed in the public sector—as the *Bonsucro EU Standard* has adapted the greenhouse gas methodology from the EU Renewable Energy Directive. While metric standards are not yet widespread, we can see an example of technology-based 'uploading' in the case of organic certification, as most organic standards started off as private initiatives before becoming more publicly governed (Bendell *et al.*, 2011), earning a degree of enforceability through the courts as a result of both legislative inclusion and private contracting (Webb & Morrison, 2004). While, under EU law, public procurement tenders cannot require a certain certificate *per se*, they can require sustainability criteria for which a certificate is 'proof of compliance'. In areas where contracting and monitoring is underdeveloped, this is a method through which the structure and requirements of private standards can enter public policy (D'Hollander & Marx, 2014). Developing a pluralism of initial methodologies through competing sustainability standards can allow different methods of measuring impacts to be evaluated and chosen by actors, including policymakers (Smith & Fischlein, 2010).

4. AN AGENDA FOR FUTURE RESEARCH

Both the field of sustainability and the practice of sustainability standards are still emerging, and there is much we do not yet know. With regards to performance-based metric standards, several areas for future research are especially striking.

With regards to flexibility of moving facilities toward sustainability, there is a surplus of theory and a dearth of empirics surrounding the on-the-ground reality of efficiency increases from performance-based standards and potential barriers to understanding and building practices around metrics. Current studies tend to be drawn from the healthcare and local government literatures (*e.g.* De Bont & Grit, 2011; Kelman & Friedman, 2009; Moynihan & Pandey, 2010), and their generalisability to a completely different context, such as a farm, is unclear. Can knowledge and practices be easily disseminated across users of the standard? Is there a role here for new communication technologies, or stakeholder methodologies? We are seeing a transnational organisational field of sustainability professionals (Dingwerth & Pattberg, 2009)—can a transnational organisational field of on-the-ground sustainable practitioners emerge?

Information provision must also prove itself empirically to be useful in a transition toward sustainable outcomes. On a facility scale, it is important to know if there are positive effects from such provision, are they significant, and are they lasting? What benefit can the information provide in practice to facilities beyond certification, and is this valued by financial markets or policy-makers? What synergies can be created between public policy and aggregate information from private schemes?



A lot of research will need to be done on the new forms that standards can take. Can we build well-founded guidelines enabling the use of decision rules that better fit the systems they are trying to govern? Is it possible to strike a balance that allows for both new rules while maintaining adequate understanding and communication by the end-users of the standards? Through what type of processes can we develop legitimate decision rules for value-laden social, economic and environmental issues? Building on work done in technology assessment [e.g. Guston & Sarewitz, 2002; Schot & Rip, 1997], how can we adequately involve stakeholders in technical matters such as decision rules and thresholds, in order to help legitimise their development? Many standards informally have reported difficulties in pushing incremental revisions—the certified user base sees it as extra compliance costs. How can a dynamic standard be created politically, given that users and stakeholders have to agree to the ‘ratcheting up’ of necessary investment?

Measurement itself is a research topic for many reasons. First, what are the role of technologies in measurement? Standard systems are just at the beginning of investigating the role that sensors and crowdsourcing can play in creating credible standards and reliable data. ISEAL Alliance, for example, has begun to scope the area with a recent research tender. Secondly, social measurement in standards is significantly more challenging than technical measurement. Social measurement encodes values, is methodologically hard, and difficult to robustly audit. Quantification may miss important dimensions of exclusion, disempowerment, discrimination and the like. Future research may wish to consider the differences between technology-based and performance-based approaches to social impact, and suggest ways forward.

Learning from data requires guidelines in order to make methods accessible for all standard-schemes. This may be less about research than sharing tools and practices between systems. More work is required to look at whether spatially heterogeneous or homogenous standards have higher impact on sustainable outcomes in the long-term, which can inform how data is used in revision processes. As with all areas in adaptive management, concrete examples and guidelines of how to strengthen organisations to allow for ‘double-loop learning’ is required in order to turn thought into practice [Argyris, 1976].

Given the novelty still surrounding metric standards and sustainability standards in general, case studies on synergies between policy instruments are still emerging, and more of those are needed to get a larger picture of how standards shape and are shaped by their institutional environments. What are the opportunities and challenges in harmonising private and public governance? In such synergies, what are the power dynamics, and how is accountability best ensured?

5. CONCLUSION

Performance-based standards open up a wide range of new opportunities for sustainability standards. Naturally, with these opportunities also come challenges and pitfalls. Metric

standards may have associated rewards that go beyond the prescription of good practices, but they also require serious investment in building relevant methods and skills. Nike, for example, spent \$6 million USD in-house on its open source *Considered Design Index* and Environmental Apparel Design Tool, as no relevant metric system was on the market and available to purchase⁵. A one-size-fits-all approach is unlikely to be possible, and standard-systems are likely to require considerable bespoke effort to fit metrics to their need, which will only be possible with capacity building in areas of data science and technical stakeholder engagement.

In summary, there is much still to be done. However, if sustainability standards began as consumer-facing labels, and are moving more and more towards business-to-business models, then the next step could very well be a move to more widely applicable and useful metrics, which can be communicated to buyers as well as be intrinsically useful to decision-makers. Certification systems must keep improving and reinventing themselves to keep up with changing socio-ecological systems and markets, and innovation and discovery in this field will be vital in their ability to enact long-term transformative change.

References

- Abbott, K.W. & D. Snidal (2009). Strengthening international regulation through transnational new governance: overcoming the orchestration deficit. *Vanderbilt Journal of Transnational Law* 42: 501-78.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics* 95(9): 1082-1095. doi: 10.1016/j.jpubeco.2011.03.003
- Ambec, S. & P. Lanoie (2008). Does it pay to be green? A systematic overview. *The Academy of Management Perspectives* 22(4): 45-62. doi: 10.5465/AMP.2008.35590353
- Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative Science Quarterly* 21(3): 363-375.
- Attig, N., S. El Ghoul, O. Guedhami & J. Suh (2013). Corporate social responsibility and credit ratings. *Journal of Business Ethics* 117(4): 679-694. doi: 10.1007/s10551-013-1714-2
- Bardach, E. & R.A. Kagan (1982). *Going by the Book: The Problem of Regulatory Unreasonableness*. Philadelphia: Temple University Press.
- Barrett, C.B., A.J. Travis & P. Dasgupta (2011). On biodiversity conservation and poverty traps. *Proceedings of the National Academy of Sciences* 108(34): 13907-13912. doi: 10.1073/pnas.1011521108

⁵ Nike Inc. Press Release, November 30, 2010: *Nike releases environmental design tool*. Retrieved from <http://nikeinc.com/news/nike-releases-environmental-design-tool-to-industry>

- Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: Sage Publications.
- Bendell, J., A. Miller & K. Wortmann (2011). Public policies for scaling corporate responsibility standards: Expanding collaborative governance for sustainable development. *Sustainability Accounting, Management and Policy Journal* 2(2): 263–293. doi: 10.1108/20408021111185411
- Berkes, F. & C. Folke (2002). Back to the future: Ecosystem dynamics and local knowledge. In: Gunderson, L.H. & C.S. Holling (Eds.), *Panarchy: Understanding Transformations in Human and Natural Systems*, pp.121–146. Washington, D.C.: Island Press.
- Bernstein, S. & B. Cashore (2007). Can non-state global governance be legitimate? An analytical framework. *Regulation & Governance*, 1(4): 347–371. doi: 10.1111/j.1748-5991.2007.00021.x
- Blann, K., S. Light & J.A. Musumeci (2003). Facing the adaptive challenge: Practitioners' insights from negotiating resource crises in Minnesota. In: Berkes, F., J. Colding & C. Folke (Eds.) *Navigating social-ecological systems: Building resilience for complexity and change*, pp.210–40. Cambridge, UK: Cambridge University Press.
- Brooks, H. (1980). Technology, evolution, and purpose. *Daedalus* 109(1): 65–81.
- Campanella, G. & R.A. Ribeiro (2011). A framework for dynamic multiple-criteria decision making. *Decision Support Systems* 52(1): 52–60. doi: 10.1016/j.dss.2011.05.003
- Carolan, M.S. (2006). Sustainable agriculture, science and the co-production of 'expert' knowledge: The value of interactional expertise. *Local Environment* 11(4): 421–431. doi: 10.1080/13549830600785571
- Carter, M.R. & C.B. Barrett (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies* 42(2): 178–199.
- Cash, D.W., W.C. Clark, F. Alcock, N.M. Dickson et al. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences* 100(14): 8086–8091. doi: 10.1073/pnas.1231332100
- Cashore, B., G. Auld, S. Bernstein & C. McDermott (2007). Can non-state governance 'ratchet up' global environmental standards? Lessons from the forest sector. *Review of European Community & International Environmental Law* 16(2): 158–172.
- Cashore, B., G. Auld & D. Newsom (2004). *Governing Through Markets: Forest Certification and the Emergence of Non-State Authority*. London: Yale University Press.
- Cialdini, R.B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science* 12(4): 105–109. doi: 10.1111/1467-8721.01242
- Clark, W.C., T.P. Tomich, M. van Noordwijk, D. Guston et al. (2011). Boundary work for sustainable development: Natural resource management at the Consultative Group on International Agricultural Research (CGIAR). *Proceedings of the National Academy of Sciences* (August 15, 2011): published online. doi: 10.1073/pnas.09002311108
- Coglianesi, C., J. Nash & T. Olmstead (2003). Performance-based regulation: Prospects and limitations in health, safety, and environmental protection. *Administrative Law Review* 55(4): 705–729.
- De Bont, A. & K. Grit (2011). Unexpected advantages of less accurate performance measurements. How simple prescription data works in a complex setting regarding the use of medications. *Public Administration* 90(2): 497–510. doi: 10.1111/j.1467-9299.2011.01959.x
- D'Hollander, D. & A. Marx (2014). Strengthening private certification systems through public regulation. *Sustainability Accounting, Management and Policy Journal* 5(1): 2–21. doi: 10.1108/sampj-04-2013-0016
- Dingwerth, K. & P. Pattberg (2009). World politics and organizational fields: The case of transnational sustainability governance. *European Journal of International Relations* 15(4): 707–743. doi: 10.1177/1354066109345056
- Faruqui, A., S. Sergici & A. Sharif (2010). The impact of informational feedback on energy consumption—a survey of the experimental evidence. *Energy* 35(4): 1598–1608. doi: 10.1016/j.energy.2009.07.042
- Fortin, E. & B. Richardson (2013). Certification schemes and the governance of land: Enforcing standards or enabling scrutiny? *Globalizations* 10(1): 141–159. doi: 10.1080/14747731.2013.760910
- Frey, B.S. & S. Meier (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review* 94(5): 1717–1722.
- Gadgil, M., N.S. Hemam & B.M. Reddy (1998). People, refugia and resilience. In: Berkes, F., C. Folke & J. Colding (Eds.) *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*, pp.30–47. Cambridge, UK: Cambridge University Press.
- Gadgil, M., P. Olsson, F. Berkes & C. Folke (2003). Exploring the role of local ecological knowledge in ecosystem management: Three case studies. In: Berkes, F., J. Colding & C. Folke (Eds.) *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change*, pp.189–209. Cambridge, UK: Cambridge University Press.
- Gerber, A.S. & T. Rogers (2009). Descriptive social norms



- and motivation to vote: Everybody's voting and so should you. *The Journal of Politics* 71(1): 178–191. doi: 10.1017/S0022381608090117
- Giddens, A. (2002). *Runaway World: How Globalisation is Reshaping Our Lives*. London: Profile Books.
- Gunderson, L.H. & C.S. Holling (2002). *Panarchy: Understanding Transformations in Human and Natural Systems*. Washington, D C: Island Press.
- Gunderson, L.H., C.S. Holling & S. Light (Eds.) (1995). *Barriers and Bridges to the Renewal of Ecosystems and Institutions*. New York: Columbia University Press.
- Gunningham, N. (1996). From compliance to best practice in OHS: The roles of specification, performance and systems-based standards. *Australian Journal of Labour Law* 9(3): 1–26.
- Guston, D.H. & D. Sarewitz (2002). Real-time technology assessment. *Technology in Society* 24: 93–109.
- Holling, C.S. & G.K. Meffe (1996). Command and control and the pathology of natural resource management. *Conservation Biology* 10(2): 328–337. doi: 10.1046/j.1523-1739.1996.10020328.x
- Inglehart, R. (1997). *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Cambridge, UK: Cambridge University Press.
- Jaffee, D. (2007). *Brewing Justice: Fair Trade Coffee, Sustainability, and Survival*. Berkeley, CA: University of California Press.
- Janssen, M.A., Bodin, J.M. Anderies, T. Elmqvist *et al.* (2006). Toward a network perspective of the study of resilience in social-ecological systems. *Ecology and Society* 11(1): 15.
- Kelman, S. & J.N. Friedman (2009). Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory* 19(4): 917–946. doi: 10.1093/jopart/mun028
- Lang, D.J., A. Wiek, M. Bergmann, M. Stauffacher *et al.* (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science* 7(S1): 25–43 doi: 10.1007/s11625-011-0149-x
- Lansing, J.S. & J.N. Kremer (1993). Emergent properties of Balinese water temple networks: Coadaptation on a rugged fitness landscape. *American Anthropologist* 95(1): 97–114. doi: 10.1525/aa.1993.95.1.02a00050
- Lebel, L. (2012). Agricultural standards and certification systems. In: The Steering Committee of the State-of-Knowledge Assessment of Standards and Certification (Ed.) *Toward Sustainability: The Roles and Limits of Certification*, pp.A125–A145. Washington, DC: RESOLVE, Inc..
- Matus, K.J.M. (2010). Standardization, certification, and labeling: A background paper for the roundtable on sustainability workshop, January 19–21, 2009. In: Committee on Certification of Sustainable Products and Services (Ed.) *Certifiably Sustainable?* pp.79–104. Washington, DC: National Academies Press.
- Mitchell, R.B., W.C. Clark, D.W. Cash & N.M. Dickson (2006). *Global environmental assessments: Information and Influence*. Cambridge, MA: MIT Press.
- Morhardt, J.E., S. Baird & K. Freeman (2002). Scoring corporate environmental and sustainability reports using GRI 2000, ISO 14031 and other criteria. *Corporate Social Responsibility and Environmental Management* 9(4): 215–233. doi: 10.1002/csr.26
- Moynihan, D.P. & S.K. Pandey (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory* 20(4): 849–866. doi: 10.1093/jopart/muq004
- Neilson, J. & B. Pritchard (2009). *Value Chain Struggles: Institutions and Governance in the Plantation Districts of South India*. Chichester, UK: Wiley-Blackwell.
- Nolan, J.M., P.W. Schultz, R.B. Cialdini, N.J. Goldstein & V. Griskevicius (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin* 34(7): 913–923. doi: 10.1177/0146167208316691
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science* 5(1): 14–37. doi: 10.1287/orsc.5.1.14
- Nussbaum, M.C. (2001). *Women and Human Development: The Capabilities Approach*. Cambridge, UK: Cambridge University Press.
- Ostrom, E. & H. Nagendra (2006). Insights on linking forests, trees, and people from the air, on the ground, and in the laboratory. *Proceedings of the National Academy of Sciences* 103(51): 19224–19231 doi: 10.1073/pnas.0607962103
- Porter, M.E. & M.R. Kramer (2011). Creating shared value: How to reinvent capitalism—and unleash a wave of innovation and growth. *Harvard Business Review* 89(1/2): 2–17.
- Powell, W.W. & S. Grodal (2006). Networks of innovators. In: Fagerberg, J., D.C. Mowery & R.R. Nelson (Eds.) *The Oxford Handbook of Innovation*, pp.56–85. Oxford, UK: Oxford University Press.
- Ribaudo, M. & M.F. Caswell (1999). Environmental regulation

in agriculture and the adoption of environmental technology. In: Casey, F., A. Schmitz, S. Swinton & D. Zilberman (Eds.) *Flexible Incentives for the Adoption of Environmental Technologies in Agriculture*, pp.7–25). Dordrecht: Springer.

Robinson, J (2004). Squaring the circle? Some thoughts on the idea of sustainable development. *Ecological Economics* 48(4): 369–384. doi: 10.1016/j.ecolecon.2003.10.017

Schot, J. & A. Rip (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change* 54(2-3): 251–268.

Schultz, P.W., J.M. Nolan, R.B. Cialdini, N.J. Goldstein & V. Griskevicius (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science* 18(5): 429–434. doi: 10.1111/j.1467-9280.2007.01917.x

Schwesinger Berlie, L. (2010). *Alliances for Sustainable Development: Business and NGO Partnerships*. Basingstoke, UK: Palgrave Macmillan.

SCSKASC [Steering Committee of the State-of-Knowledge Assessment of Standards and Certification] (2012). *Toward Sustainability: The Roles and Limitations of Certification*. Washington, DC: RESOLVE, Inc..

Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Smith, T.M. & M. Fischlein (2010). Rival private governance networks: Competing to define the rules of sustainability performance. *Global Environmental Change* 20(3): 511–522.

Tallberg, J. (2002). Paths to compliance: Enforcement, management, and the European Union. *International Organization* 56(3): 609–643. doi: 10.1162/002081802760199908

UNCTAD [United Nations Conference on Trade and Development] (2011). *World Investment Report 2011*. Geneva: United Nations.

Veleva, V., M. Hart, T. Greiner & C. Crumbley (2001). Indicators of sustainable production. *Journal of Cleaner Production* 9(5): 447–452. doi: [http://dx.doi.org/10.1016/S0959-6526\(01\)00004-X](http://dx.doi.org/10.1016/S0959-6526(01)00004-X)

van der Wal, S. (2011). *Certified Unilever Tea: Small Cup, Big Difference?* Amsterdam: SOMO.

Walters, C.J. (1997). Challenges in adaptive management of riparian and coastal ecosystems. *Conservation Ecology* 1(2): 1.

Webb, K. & A. Morrison (2004). The law and voluntary codes: Examining the 'tangled web'. In: Webb, K. (Ed.) *Voluntary Codes: Private Governance, the Public Interest and Innovation*, pp.97–104. Ottawa: Carleton Research Unit for Innovation, Science and Environment, Carleton University.