312 Clay Street, Suite 300
Oakland, CA, 94607-3510 USA
tel.: +1-510-839-8879
info@footprintnetwork.org
www.footprintnetwork.org
contact: William Coleman

Comments on the Industrial Footprint Project Tool and Reports

INTRODUCTION

Global Footprint Network highly appreciates the opportunity to comment on this important initiative. The goal of this initiative is of stellar importance – and using management tools for advancing sustainable economies is critical.

Having read the reports, we see the biggest gaps in the framing of how the project is being approached. Therefore, rather than providing detailed feedback on the particular scoring card or *Industrial Footprint Tool*[1] developed by the Washington State Department of Ecology and Earth Economics for the pulp and paper industry, Global Footprint Network's comments address the more fundamental issue of the problematic nature of scorecards as metrics, and why accounting approaches are needed rather than arbitrary score cards.

We suggest a process for revision or extension of the project:

- Identify the ultimate goals you want to achieve;
- Translate these goals into empirically based research questions (these become your outcome indicators);
- Specify your theory of change for achieving your ultimate goals and detail your key interventions (from which you can develop process indicators).

In other words, we would argue that two distinct types of metrics are needed: outcome indicators that track the ultimate goal of a sustainable, prosperous and equitable economy, and process metrics that evaluate the effectiveness of specific intermediate interventions.

SCORECARDS VS. ACCOUNTS

Scorecards, as proposed in the reviewed report, do not have a clear scientific basis since they lack specific, empirically based research question.[2] Without having clarity about the question,

[1] Please also note that in naming of the scoring system, using the term "Footprint" may be misleading since the metric does not compare resource demand to biocapacity. It may therefore be more appropriate to call the metric a "sustainability scoring system."
[2] The term *scorecard* refers to the aggregate, synthetic summary representation of the individual measures. Individual measures (such as pH level of water, or carbon emissions, or percentage of renewable energy used on premises etc.) are empirical measures and within the realm of science. It is the aggregation that leaves the realm of

such approaches produce results that are ultimately uninterpretable. As arbitrarily aggregated indices, they cannot show tradeoffs. Their value, rather, is that they serve as conversation tools or "alarm bells" when a system, in aggregate, moves out of the allowable range.

We suggest that the reviewed report reframe the presented data as a *database* or *key indicators set* describing aspects of the mill. This makes clear that the indicators are a collection of data, rather than a framework.[3] The name "Footprint" suggests a framework, a comprehensive account. In addition to the concerns expressed above, there may also be a risk that by focusing too early on data or indicator collection, the project may lose sight of what the key questions (or information) are for operating a sustainable operation.

We recognize that scorecards are used often in public policy or sustainability reports, even by reputable universities. But this does not change the fact that they are unscientific by nature.

Compare the situation to flying a plane. The pilot needs an airplane dashboard that gives her good answers to the most relevant (and clearly stated) questions that she needs to keep tack of in order to keep the plane safe and oriented. There are several relevant questions a pilot needs to track (altitude, speed, amount of fuel left, direction). But mixing various clear questions into an overall index destroys relevant information. No airplane would be flown on an aggregate index of altitude, cabin pressure, fuel left in tank, and speed – all need separate metrics. There is a minimal set of metrics needed to fly the plane safely and comfortably. Establishing what this set is, and how these measures relate to each other, is essential for the pilot.

Dashboards, as collections of discrete indicators, are designed by asking first what the critical information is that a pilot needs. These most critical pieces of information are placed at the centre for the pilot's ease. This means that dashboards are more than just a collection of data. They provide a structured framing to help identify the relevant information and make sense out of it.

Mill operators, likewise, need metrics that start from a clear, empirical research question (which is missing in the Industrial Footprint study). Clear questions produce clear, empirically based metrics. To the extent that our metrics are grounded in clear research questions and do not rely on arbitrarily aggregated indices, we are in the realm of science. Hence the mill project would gain from adopting an aeronautical engineering approach, and determine what an ideal dashboard would need to look like.

---

natural science since the weighing is arbitrary. This does not mean that scorecards have no value. If designed carefully, the can be alarm-bells. Or if used in a consensus process they can reveal priorities of the people involved in the census process. A mill can improve each single indicator, but the arbitrary weighing will not get an adequate insight on where they should put there efforts.

[3] It is useful that the indicators are organized in categories (in your case - economic, social ecological), but with framework we mean an explicit system of drivers and outcomes of a system. These are discovered by finding the key questions needed for being able to manage the system.

Figure 1 sketches a continuum extending from broad scorecards with no theoretical or scientific basis to the most rigorous scientific accounting tools. We would recommend adopting metrics that fall as far as possible to the right-hand side of this continuum, while still capturing the breadth of the question that needs to be covered (figure 1).

Figure 1: Continuum from scorecards to accounts (with examples)

Theory-less scoring systems: **quiz in magazine**	Index of a broad theme, mixing domains: **EPI, SPI, pulp score**	Valuing various dimensions of a singular phenomenon: **HDI, EMC**	Relative Index of one dimension: **LPI, Dow Jones**	Accounts based on a research question: **GHG, MFA, HANPP, Footprint, GDP**	Counting of the same units: **carbon, land-use areas**

Broad, but more arbitrary scorecards ⟷ Narrow, but science-based accounting

Accounts are preferable to indices insofar as they are based on scientific principles and can show trade-offs among the parameters covered. Indices and scorecards function more as alarms: They are sensitive to alarming levels within any of the covered domains, but cannot accurately describe trade-offs among the domains since the weighting is based on arbitrary choices (note: even equal weighting is a kind of weighting). Here the airplane analogy comes in handy – an alarm goes off if any of a number of distinct parameters are out of safety range, but the alarm might mean one of many things: lack of fuel, or dangerously low altitude, speed, or cabin pressure. The alarm's primary function is to get the pilot's attention; she will need more specific information to find out exactly what the problem is.

To operate, mills will have to make decisions about where to put their efforts. Being faced with many indicators in the data set, the question becomes which ones to pick. Obviously, if some indicators were outside a legal range, they would get priority, but what about the others? And might there be blind spots? These perspectives emphasize the need for clear essential questions that drive the inquiry.

Recommended Process for Revising Project

Identify ultimate goals

What is the ultimate goal of our endeavor? Many would say: sustainable development. This implies specific outcomes. On the one hand, we want to achieve *development*, which is shorthand for great, fulfilling lives for all human beings. On the other hand we want it to be *sustainable*, which means fitting within the regenerative capacity of the planet.[4] The ultimate goals, or to use a term from analytic philosophy, *ultimate end*s, are fulfilling lives for all human beings. The resources at our disposal for bringing this about, i.e. the *ultimate means*, are the Earth's biocapacity. Sustainable development – great lives for all within the ecological limits of our planet – addresses both *ultimate means* and *ultimate ends*.

Formulate a clear research question
The first step toward developing a science-based method is to formulate a clear research question (a question that becomes answerable through empirical research, in the sense used by Karl Popper). Identify the question you want to answer, and frame it as a research question, with all elements being empirically defined, allowing the results to be clearly measurable.

Develop outcome metrics based on ultimate goals
A distinction must be drawn between outcome metrics, which measure performance based on an ultimate goal (e.g. long, happy lives for all), and process metrics, which measure the impact of specific intermediate interventions. Once the research question has been defined, the next step is to identify outcome metrics that specifically evaluate performance toward the stated ultimate ends.

<u>Table 1: Examples of outcome measures for the mills</u>

Examples of outcome measures on ultimate means	Examples of outcome measures on ultimate ends
• Biodiversity • Sustainability of source material • Resources consumed/resources renewed by region • Initial Puget Sound Partnership goals: impeccable water quality	• Employees: personal satisfaction • Health and safety: injury rate, incidents, loss time • Job satisfaction

Describe your theory of change, identify key interventions and process metrics
Ultimate ends and outcome metrics having been identified, we would recommend mapping out your organization's theory of change: How will the ultimate ends be accomplished? What are the precise mechanisms (technological, societal, etc.) by which transformation can occur? You may find it useful to represent the system in the form of a diagram.

[4] This section builds on M. Wackernagel and D. Deumling, "Future-Friendliness in Action: A Toolkit to Help Prepare the Quantum Leap towards Sustainability", in D.J. Rapport, W.L. Lasley, D.E. Rolston, N.O. Nielsen, C.O. Qualset, and A.B. Damania, ed., *Managing for Healthy Ecosystems* (Boca Raton: Lewis, 2003). The paper inspired by Donella Meadows' and Herman Daly's 'pyramid'. See D. Meadows, *Indicators and Information Systems for Sustainable Development* (Hartland Four Corners: Sustainability Institute, 1998): 40-47; see also Herman Daly, ed., *Toward a Steady-State Economy* (San Francisco: W.H. Freeman, 1973): 8. Daly defines ultimate means as the "fundamentally useful stuff of the universe, i.e., low-entropy matter-energy." Global Footprint Network would call it "biocapacity".

Your theory of change will most likely include specific interventions geared toward effecting specific types of change. After mapping out your theory of change, the next step is to develop process metrics to measure the efficacy on each planned intervention.

REFINING KEY CONCEPTS: "ENVIRONMENTAL IMPACT"

One further limitation of the Industrial Footprint project is that it conflates the ultimate means – what the Earth is able to provide – with "environmental impact". Environmental impact is an aggregate of various subsidiary concepts, an umbrella idea – not one specific, measurable thing. It includes an array of factors, possibly best characterized on a two-pole continuum between biocapacity-related concerns (greenhouse gas emissions, food and timber production, fishery collapse, etc.) and human health-related concerns (air and water quality, toxics, etc.).

Both biocapacity concerns and human health-related concerns are important, but they show quite distinct characteristics (see Table 2), and in fact the first is part of ultimate means (access to sufficient biocapacity), the second part of ultimate ends (healthy lives). Together the two concerns cover a wide range of what we would consider "environmental impacts" (Table 3). Aggregating the two is confusing, as they operate differently and need to be looked at with separate tools. Traditionally, the human health dimension has been covered more effectively. But robust indicators are needed for each domain.

Accounting approaches are possible for both domains. For instance, the Ecological Footprint is an accounting tool answering the specific question: How much biocapacity is needed to support specific activities? For example, the human health side can be captured in measures such as life-years are lost due to specific emissions.[5] We are not arguing that this should be done for this particular project. Assessing potential lost life-years is challenging. But by posing the question, the intent becomes clear. Even without having the measure, the question helps to sharpen the discussion about trade-offs.

Table 2: Characteristics of the two domains of environmental concerns

Biocapacity Concerns (Resource demand vs. biocapacity)	**Human Health Concerns**
• Global scale • Quantitative problem • Slow improvements • Use of resources is essential to our	• Side-effect of main activity • Local phenomenon • Qualitative problem

[5] Eco indicators 99 use this vantage point to aggregate the impact on human health. (http://www.pre.nl/eco-indicator99/eco-indicator_99.htm). Even thought such assessments might be difficult and costly, and potentially unpractical or financially prohibitive, we still consider it more productive to be clear about the underlying intent (i.e., what we would like to measure ideally), rather than to aggregate arbitrarily.

Metabolism, therefore problem needs broader policies than merely technology • Decoupling more challenging. As biocapacity limits are breached, impact per unit of demand increases (for example, fishing in overfished ocean, emitting CO_2 in a more carbon-constrained world, removing freshwater in a drought)	• Quick remediation • Typically substitutes are available, or can be found; therefore, problem is technologically solvable • Decoupling possible

<u>Table 3: Examples of impact areas covered by the two domains</u>

Biocapacity	**Human health**
• Resource consumption • Land use • Climate change • Stratospheric ozone depletion • Acidification • Eutrophication • Impact on ecosystems and biological diversity	• Human health impacts • Eco-toxicity • Photo-oxidant formation • Ionizing radiation

CONCLUSIONS: BUILDING STAKEHOLDER ENGAGEMENT

Finally, progress is not possible without broad stakeholder engagement. Therefore we are glad to see careful attention given to that process – as documented in the “Stakeholder Engagement Report”.

It is great to see the highly interactive workshops being held. We would suggest focusing later on specific indicators, and initially more on the questions stakeholders have, and information needs they have in order to trust the operation.

The Engagement report gives the following agenda for the first workshops:

1. General introduction and project overview presented by Department of Ecology representatives.
2. Presentation by mill representatives on project participation.
3. Small-group brainstorming sessions in which participants were given the opportunity to generate potential indicators for each domain (environmental, social, economic).
4. Full-group review of indicator brainstorm results.
5. Large-group preference rating process in which all participants were provided with 3 colored dot stickers to place on the indicators that held the greatest importance from their perspective on development of final tool indicator sets. Rating preference data were collected for general reference in subsequent stages of indicator selection and weighting, and it was explicitly stated that the preference rating process did not in any way represent a voting process.
6. Summary of progress and next steps.

We would have suggested an agenda along the following lines:

1. General introduction and project overview presented by Department of Ecology representatives.
2. Presentation by mill representatives on project participation.
3. Small-group brainstorming sessions on the following questions: what are the ultimate goals of this endeavor? What do you want as a community? What is the information they need in order to a) operate the mill, b) trust the mill (know that it is a positive force for the community). This needs to be guided by good facilitation to help them identify the big principles and goals, rather than getting lost in details. Details and specifics are important to identify larger principles. I would expect questions like: are we running out of resources? Are the effluents affecting our health? Is the mill becoming economically more successful or less successful? Is it providing satisfying, safe and well-paying jobs to community members? Does the community have a say in the mill’s operation? Is there a danger of the mill’s being pulled out of the unexpectedly due to decisions at headquarters? Etc.
4. Full group review of key principles and needs for community (and mill).

5. Large-group preference rating process in which all participants were provided with 3 colored dot stickers to place on issues that held the greatest importance from their perspective.
6. Once the principles are clear, only then start discussion on the question of how to measure whether we fulfill these principles or not.
7. Summary of progress and next steps.

In essence, in order to build this support, the process has to start with exploring what the relevant questions are. Some say, “Answers kill inquiries”. Coming first with answers polarizes and makes people defensive. Everybody is defending their answer – nobody asks what questions these answers are supposedly addressing.

In summary:

a) Be clear what your ultimate goal is, and build stakeholder buy-in to a shared vision of the future;

b) Identify the critical question (or information need) to support these goals, and formulate it as an empirically-based research question which then in turn define your ***outcome indicators***;

c) Make clear your theory of change;

d) This theory of change identifies milestones along the way of producing the desired outcomes. These measurable milestones then become your science-based ***process indicators***.

The UNIVERSITY of VERMONT

DEPARTMENT OF COMMUNITY DEVELOPMENT
AND APPLIED ECONOMICS

Introduction:

Indicators for tracking the ecological, social and economic sustainability in the pulp and paper industry, and in all other industries for that matter, are sorely needed. My understanding of tracking sustainability includes:

- The overall goal is to increase mill and industry sustainability.
- An indicator should be evaluated according to how effective the overall approach is in achieving this goal.
- We must assess whether it includes the proper measures to communicate to the target audience.
- We need to make sure the measures are well designed, robust across mills and resilient to environmental, social, economic and technological changes.

I will evaluate the Industrial Footprint Project (IFP) according to the first and third measure, as I do not have adequate knowledge of the pulp and paper industry to evaluate the second. Unfortunately, I have not had time to carefully read the section on stakeholder involvement. How stakeholders are involved depends on the target audience, discussed below. It is important to note I have not had time to do a thorough evaluation of this project, but hope that some of my general comments prove useful.

Overall approach:

My first comment is that comparability between mills should be a goal of the project. Since mills are different sizes, comparisons need to be relative, e.g. environmental impacts per million dollars gross revenue, per million dollars profit, per job created, per million dollars in tax revenue, etc. Lack of this applicability is perhaps the greatest weakness of the project.

To explain, a sustainability indicator can be targeted towards different audiences. Which audience is best to target depends on our assumptions concerning mills' motivations for becoming more sustainable, and also determines in part how the indicator should be designed. An indicator can target consumers so that they will be able to compare the sustainability of different mills and make appropriate choices. If consumers prefer to buy from more sustainable mills, it will put pressure on mills to become more sustainable. It can be targeted towards regulators, to inform them of the impacts of mills on society, of the best practices available to mills, and potentially of the trade offs between different types of sustainability. This helps regulators develop appropriate policies for promoting sustainability, and assumes that mills are

College of Agriculture & Life Sciences, Morrill Hall, 146 University Place, Burlington, VT 05405-0106
Telephone (802) 656-2001, FAX (802) 656-1423, http://www.uvm.edu/~cdae/dept/

best motivated by legal restrictions. It can be targeted towards the mills themselves to give them the information they require to improve their performance. This is the only circumstance where it would not be necessary for the indicator to allow comparisons between mills. It also assumes that the mills have some intrinsic motive for becoming more sustainable, while the most likely motive is actually pressure from consumers or government regulators. If so, then the consumers and government should also be part of the target audience, and the indicator should allow comparison between mills.

The fact that this indicator does not allow comparison between mills implies that it is primarily targeted towards the individual mills themselves. I believe that the intrinsic motivation of mills is likely to be the least important factor promoting mill sustainability. In fact, if the sustainability scores are made available to consumers, they are likely to interpret them as relative measures of sustainability across mills, rather than measures of sustainability across time within a single mill. It is in fact misleading if they are not. We know from experience that when the government demanded that mills publish their annual toxic emissions, the most polluting mills made substantial efforts to reduce emissions, even in the absence of any regulation. No one wanted to one of the 10 worst polluting mills in the country. Since someone is always one of the 10 worst mills, simply making information on toxic emissions public has proven a steady source of pressure for reducing emissions.

Why footprint?

Another basic problem with the IFP is that it does not actually try to estimate a footprint, so its name should be changed. By definition, an ecological footprint tries to estimate how much biophysical productive capacity is utilized or degraded by a person, mill, industry or political region. This is calculated in terms of hectares, a physical unit of area. The analogy is with the physical space covered by an actual footprint. How much biological productivity are we stepping on? I'm not even sure how to think about a social footprint or economic footprint—how much social production is the mill stepping on? It makes no sense to me. You're really talking about sustainability measurement, which should not be called a footprint at all. Additionally, though impossible to achieve, the ideal goal for an ecological footprint is zero. The IFP instead strives for a maximum score of 100. My overall impression is the name of the project was decided on first. Then a decision was made not to alienate the worst polluting industries by refusing to allow comparability.

Designing the measures

Of the five approaches to measuring indicators, comparison with a base year is the least appropriate. Using the footprint analogy, it implies that shoe size is determined by the percentage growth of your foot in a given year. As people get older, percentage growth in shoe size declines, and as we mature, we all have a score of zero (good for an ecological footprint, bad for the measure you're discussing here). In the case of the IFP score, you would always have to use the same baseline year. If you change baseline years, you could actually see scores getting worse and worse as mills improved their practices over time. If you don't change baseline years, you can't apply the scorecard to new mills, which makes it not at all robust, and you give a permanent advantage in scores to the worst performing mills. As a measure of a footprint, this score is not simply weak, but actually perverse.

College of Agriculture & Life Sciences, Morrill Hall, 146 University Place, Burlington, VT 05405-0106
Telephone (802) 656-2001, FAX (802) 656-1423, http://www.uvm.edu/~cdae/dept/

The other four measures also fail to measure footprints, though they seem to be potentially appropriate measures of environmental performance at a point in time. Measures related to best and worst practices however are not particularly robust over time, because acceptable standards and best available technologies ideally change over time. With technological advances, all mills should eventually exceed targets, and have perfect scores, so the measure becomes useless. Perfect scores for a given indicator are only appropriate if the target implies a zero footprint. For example, if targeted SO_x emissions are set equal to the absorption capacity in the areas over which they precipitate, which would be determined by the rate at which buffers (e.g. calcium carbonates) leach into the soil. Even then, a perfect score is only appropriate if no one else is emitting SO_x now, no one else would like to emit SO_x in the future, and there is no environmental damage from emissions below waste absorption capacity. In this case, there is no externality or opportunity cost associated with a mill's emissions, and no marginal benefit to reducing them.

The ecological footprint measure would convert all throughput (i.e. resource inputs (raw materials) and waste outputs (pollution)) from a given activity into an area of biologically productive land required to sustain that activity. I suggest an appropriate measure of a footprint should simply be the area of biologically active land utilized by the industry[1]. This can be given as a raw score or standardized, (e.g. emissions per air dried ton) to facilitate comparison between mills of different size. Using the example of SO_x, Boise emits 2.9013 lbs per ADT. To translate this into an area, in principle one would have to estimate how much biologically active land is required to neutralize this quantity of acid, perhaps 15 hectares per year (I actually have no idea what the real number is, and am assuming that there is an annual influx of acid buffering compounds into the soil). In this case, the SO_x footprint is 15 hectares. Of course, land that sequesters SO_x can also process other wastes simultaneously, and generate resource inputs as well. I do not know how the ecological footprint addresses this issue, but I would suggest assigning a weight to each component of the footprint such that all weights sum to unity then calculating the total footprint. However, I believe that it would much easier simply to abandon the footprint language.

Unnecessarily complicated explanations confuse the issue

Explanations for scoring related to best and worst industry practices are unnecessarily complicated and confusing. Whether you are trying to minimize or maximize your indicator, all you need to do is determine in percentage terms where the raw indicator value falls between the extremes. The very simple equation is 100*(indicator score – industry worst)/(industry best – industry worst). If the indicator score falls between the best and worst, this number will always be positive. If the industry does worse than the worst allowed, this number will be negative, but why not penalize a mill for falling below the lowest acceptable level? If the mill exceeds the best standard, the number will exceed 100, but again I see no reason why the mill's excellence should not be recognized, except under the conditions I described above.

[1] I believe that the ecological footprint breaks the planet down into 9 types of productive zones, estimates the biologically productivity of each per hectare, then weights the area of each zone by its productivity. These are then summed to get the productivity of the planet, and normalized so that the summed weighted areas are equal in size to the Earth's land area.

College of Agriculture & Life Sciences, Morrill Hall, 146 University Place, Burlington, VT 05405-0106
Telephone (802) 656-2001, FAX (802) 656-1423, http://www.uvm.edu/~cdae/dept/

Weighting

The description of how individual indicators are weighted is unnecessarily complex and inflexible. Why not allow any weight, then use the spreadsheet to standardize? All you need to do is assign a weight for each indicator, say a number between 0-10 or even between 0-100. You then multiply the indicator by the weight then divide by the sum of all weights assigned.

Standardization

Finally, since you are dealing with different size mills with different levels of output, there is a need to standardize. For environmental measures, indicators are standardizing to air dried tons. However, I assume some mills produce very low priced newsprint, while others might produce very high quality paper, in which case standardizing to air-dried tons loses a lot of detail. Another option would be to standardize to gross revenue, annual profit, or some other economic metric. Each has its own strengths and weaknesses. You standardize some economic data to employee. It seems arbitrary to standardize to different metrics.

Joshua Farley, PhD
Associate Prof, Community Development and Applied Economics
Fellow, Gund Institute for Ecological Economics
205 B Morrill Hall
University of Vermont
Burlington, VT 05405
Phone: 802-656-2989
Fax: 802-656-1423