SYNTHESIZING AND STANDARDIZING CRITERIA FOR THE EVALUATION OF SUSTAINABILITY INDICATORS IN THE WATER SECTOR

Pires, A. [1*]; Morato, J.[2]; Peixoto, H.[1]; Bradley, S.[1]; Muller, A.[3]

ABSTRACT

Indicators are one of the tools available in planning and management projects that aid in the decision-making process and the monitoring of those decisions on the path towards the sustainable use and management of natural resources. However, the quality and reliability of the indicators depends on the constant improvement of the means to assess and design criteria sets. The identification and selection of criteria to evaluate indicators is not a trivial task. The research identified a proliferation of unconsolidated criteria in use in the sustainability and water resource management domains. In response, a process of synthesis and consolidation was undertaken in order to reduce the level of redundancies and to identify possible candidates for "core criteria" that are identified as being a relevant part of most evaluation frameworks. A representative collection of sources from the specialized literature was screened for evaluation criteria. In total, 74 sources were examined, containing a total of 346 mentions of criteria used for indicator assessment. An in-depth synthesis was performed using a structured matrix to organize and identify the redundancies in the criteria being utilized. The analysis permitted a reduction of the 346 criteria found to 60 unique criteria. The study proposes a standard name and a description for each criterion, aiming to provide more clarity and reduce ambiguity. The criteria were also ranked to identify which criteria were in more systemic use. Of the 60 criteria found, the 12 most cited were identified as possible core criteria for framework development. Also, in order to facilitate the design of indicator sets, all 60 criteria were divided into two approaches (scientific/top-down or end-use/bottom-up). This study identified significant redundancies and a lack of standardization in the use of criteria and it also ranked criteria to facilitate multi-method framework development. Thus, it is crucial that indicator developers not only consider criteria that have some level of standardization to be able to compare and communicate with other agencies and communities but also consider how to utilize core-criteria in the design of indicator sets.

Keywords: indicator; criteria; selection; top-down and bottom-up approach, synthesis, standardization.

[1] Universidade Federal da Bahia
[2] Universidad Politecnica de Catalunya
[3] International Living Future Institute
[*] Corresponding author: alex.pires@ufba.br/ +5571991780250

SYNTHESIZING AND STANDARDIZING CRITERIA FOR THE EVALUATION OF SUSTAINABILITY INDICATORS IN THE WATER SECTOR

ABSTRACT

Indicators are one of the tools available in planning and management projects that aid in the decision-making process and the monitoring of those decisions on the path towards the sustainable use and management of natural resources. However, the quality and reliability of the indicators depends on the constant improvement of the means to assess and design criteria sets. The identification and selection of criteria to evaluate indicators is not a trivial task. The research identified a proliferation of unconsolidated criteria in use in the sustainability and water resource management domains. In response, a process of synthesis and consolidation was undertaken in order to reduce the level of redundancies and to identify possible candidates for "core criteria" that are identified as being a relevant part of most evaluation frameworks. A representative collection of sources from the specialized literature was screened for evaluation criteria. In total, 74 sources were examined, containing a total of 346 mentions of criteria used for indicator assessment. An in-depth synthesis was performed using a structured matrix to organize and identify the redundancies in the criteria being utilized. The analysis permitted a reduction of the 346 criteria found to 60 unique criteria. The study proposes a standard name and a description for each criterion, aiming to provide more clarity and reduce ambiguity. The criteria were also ranked to identify which criteria were in more systemic use. Of the 60 criteria found, the 12 most cited were identified as possible core criteria for framework development. Also, in order to facilitate the design of indicator sets, all 60 criteria were divided into two approaches (scientific/top-down or end-use/bottom-up). This study identified significant redundancies and a lack of standardization in the use of criteria and it also ranked criteria to facilitate multi-method framework development. Thus, it is crucial that indicator developers not only consider criteria that have some level of standardization to be able to compare and communicate with other agencies and communities but also consider how to utilize core-criteria in the design of indicator sets.

Keywords: indicator; criteria; selection; top-down and bottom-up approach, synthesis, standardization.

1. Introduction

Indicators are relevant instruments to measure, communicate, and organize environmental information (Heink & Kowarik, 2010; Lyytimäki, et al., 2013). Nicholson et al (2012) show that sustainability indicators can be powerful policy decision tools when well-integrated into the management process. Therefore, they should present attributes considered relevant by the different groups, police makers, researchers, etc. that develop and apply them (Klug & Kmoch, 2014).

Indicators are an integral part of a cyclical management process as they support the decision-making process as well as enable society to monitor the progress of the policies and programs utilized to move towards higher levels of sustainability (WWAP, 2003). Work on developing sustainability indicators has improved significantly since the 1980s and the need for more consistent and applicable indicator systems continues to be evident (Spreng & Wils, 2000; Becker, 2010). Juwana, et al., 2012; Spangenberg, 2008; McCool & Stankey, 2004, among others confirm that the extensive and intense application of indicators by a wide range of users in different contexts is leading to vast array of approaches, methodologies and terminologies, often with much overlap and redundancy. If indicator use is to remain effective, there is a need to cycle between moments of creativity in which new indicators and methodologies are proposed and developed, and moments of analysis, synthesis, and standardization in order reflect on the needs of sharing information and to avoid the creation of an overwhelming number of indicators with high levels of redundancy which may damage the effectiveness of indicator use overall.

2. INDICATOR EVALUATION METHODS

There is often a tension between the need to produce scientifically verifiable and relatively standardized indicators that allow regions to be compared versus the use of community developed creative and local-specific indicators (Bell and Morse, 2003; Reed et al., 2005). Indicators often need to be compared or shared by regions or inserted into the similar project management processes of international development agencies. Thus, the verification and consolidation of indicators is becoming a significant area of research and improvement.

There are several methods that can be used check the quality of an indicator or a set of indicators, each of which will contribute differently to consolidating phases of indicator development (Aveline et. al, 2009; Bockstaller & Girardin, 2003; James et al., 2012; Cloquell-Ballester et al., 2006). These include:

- Modeling.
- Expert evaluation.
- Referential criteria.

- Participatory selection.
- 3S method - self, scientific and social validation.
- Similar real data comparison.
- Methodological revision.
- Application and observation (usefulness test).

This article focuses on the consolidation of referential criteria as used in the sustainability/water resource nexus. The use of criteria is the most common method of evaluating and validating indicators and is therefore a critical area for research. Interesting studies into the combination of evaluation methods is undergoing but as long as criteria-based evaluation continues to be important, no sector can completely avoid the need for periodic revision and consolidation.

3. CRITERIA EVALUATION

To contextualize the notion of criteria, it is common to consider that there are overarching general principles of which criteria are a more specific positive characteristic and finally the indicator are the measures themselves that synthesize agent relationships and "enable us to gain an understanding of a complex system [...] so that effective management decisions can be taken that lead towards initial objectives" (Mitchell et al., 1995). Mendoza & Prabhu (adapted 2000) present this hierarchy as follows:

- Principle: A fundamental truth or law as a basis for reasoning or action that provides the justification for the criteria as verifiers.
- Criteria: Characteristics or standards that convey specific meaning and/or operationality to a principle without itself being a direct measure of performance.
- Indicator: A measure of an agent relationship or support system performance used to infer attributes of the success or sustainability of the relationship.

Criterion are, therefore, most often utilized as the means of describing the characteristics that will assure effectiveness in indicators or indicators sets though such characteristics are relative and respective to the users of those indicators. The total number of existing criteria appears to be in the order of hundreds and still growing (WWAP, 2006; UNEP, 2006; and Niemeijer & Groot, 2008).

The Bellagio Principles (Hardi & Zdan, 1997) were an initial attempt and making a connection between principles and criteria in order to aid in the development of criteria sets and indicator choice. It was an important first step in the process of consolidation of criteria and they reinforce the notion that the identification of core criteria may work as an initial guide for the development and use of indicators. The principles serve as a general guide to aid in the process of synthesizing and integrating the work of different groups producing indicators world-wide.

The quality and reliability of the indicators can be directly affected by the application method and the appropriate choice of the criteria used to assess them (Niemeijer & de Groot, 2008). These authors, Gudmundsson (2010) and others argue that the identification and selection of criteria to evaluate indicators is not a trivial task. The process of selection should be done in a transparent and scientifically-valid way and it should incorporate significant stakeholder participation but not at the expense of becoming an overly unstructured process, as a balance is needed between top-down and bottom-up approaches (Reed et al., 2005).

Not only is it necessary to work on the improvement of the indicators themselves but it is also necessary to evaluate and consolidate the quality of the criteria being used. A significant number of works and studies have been published with this goal in mind (Gudmundsson, 2010; Kurka & Blackwood, 2013). The review of criteria-based methods done by Gudmundsson (2010) demonstrated that "a rich palette of criteria – more or less well-defined - is available to pick from the literature, but a universal list of criteria for assessing indicators does not exist." Furthermore, there is still no scientific consensus regarding which criteria are considered most relevant, although many criteria appear multiple times across a wide range of studies and may soon gain the status of core criteria for most applications.

Although there are many methodologies that consider the criteria selection process and quality of the criteria themselves (Castillo & Pitfield, 2010; Calliera et al., 2013; Khadka, & Vacik, 2012; Tanguay et al., 2013; Mascarenhas et al., 2015; Breslow et al., 2017), they present more similarities than differences. Currently most include some component of multi-stakeholder consultations and can be considered hybrid expert/participative approaches. Some form of weighting or MCA (Multi-Criteria Analysis) systems are typically applied. This study also favors a hybrid approach as the most valid. A combination of the following methods is considered a robust form of designing a framework, but this article will restrict itself essentially to the comparative study method as a necessary first stage before more complex combinatorial methods can be explored for the sustainability/ water resource nexus:

- Expert and peer review methods/ top-down approaches.
- Comparative studies.
- Framework selection.
- Bottom-up approaches.

All of these methods can contribute to the improvement of criteria selection in different manners. This study will mention each of these methods, but present results focused on a needed review and consolidation of criteria use in the water resource sector. Some implications for further combination of these evaluation methodologies will be touched upon but only as needed to contextualize the process of synthesis and consolidation.

Top-down expert and peer review methods will tend to present logical consistencies and cover more complex themes and processes. Decision makers will often automatically look to those with professional and technical experience to evaluate and suggest criteria sets. Structured expert consultations with techniques like the Delphi method can be utilized with success (Khadka, & Vacik, 2012). Statistical methods such as cluster analysis, and principal components analysis can also play an important part. The key quality that top-down methods can bring is the assurance that the indicator set is designed in such a manner to account for most of the observed changes and can therefore supply decision makers with the base information needed (Reed et al., 2005: 411). There are cases where expert consensus does not lead to methodological clarity, but some significant component of expert structuring is usually considered an important phase. In sum, top-down approaches have a place in the criterion evaluation processes and will lead to more scientifically consistent criterion sets.

Comparative methods use classification research methodology to consider lists of criteria already in use and are especially useful in identifying redundancies and increasing standardization (Deming & Swaffield, 2011). It can also help identify gaps in knowledge. As mentioned above, it is important that both criteria and indicator development pass through phases of creation and standardization so that organizations and regions can work together or compare data on management processes. Comparative methods will be an important part of the early stage of most evaluations and it is this method that is under focus here.

An interesting avenue of criteria evaluation are those that use a type of framework analysis to develop indicator sets. In this approach, the indicator set is the key aspect and not the quality of the individual indicator (Niemeijer, 2008). Or as Dahl (2012) states: "Simply compiling many separate indicators of sustainability cannot provide an adequate measure of the overall sustainability of the system." Again, there are a wide range of different frameworks used to help in the selection and modification of indicators. Historically, variations of the DPSIR (Driving Force– Pressure–State– Impact-Response) or the cause-effect approach has been a principal system used to structure the selection process and suggest cohesive indicator sets (WWAP, 2003). The DPSIR model like most frameworks help to spread the indicators over a range of human and natural processes, in theory, giving the overall evaluation more balance.

Niemeijer (2008) suggests that causal-effect frameworks can be improved to have more effect on the formation of the indicator sets instead of being used as just an organizational method. The causal-effect framework is also considered as a type of systems approach which considers the analysis of system inflows, stocks, and outputs, according to the concept of system dynamics (WWAP, 2003).

Other common frameworks include the compartment approach in which the criterion cover a range of compartments such as water, air, earth, and biota, or similarly a logical framework where the framework is structured by goals which lead to activities. For example, the goal of poverty reduction is realized by capacity building in agricultural systems and the effectiveness of the capacity building process will need to be measured by indicators. Kurtz et al., (2001) take a slightly different approach offering a framework that presents criteria in a hierarchical organization that consider four principal groups of criteria for indicator selection and development. These are:

- Conceptual relevance
    - Relevance to the assessment
    - Relevance to the ecological resource or function at risk
- Feasibility of implementation
    - Feasibility of data collection methods
    - Feasibility of the logistics
    - Feasibility of the information management
    - Feasibility of the quality assurance
    - Feasibility of the monetary costs
- Response variability
    - Estimation of measurement error
    - Within-season temporal variability
    - Across-year variability
    - Spatial variability
- Interpretation and utility
    - Discriminatory ability
    - Data quality objectives
    - Assessment thresholds (for when to take action)
    - Linkage to management actions

With any framework, the need is for overriding structure, balance, and a means of maintaining a wholistic approach when considering the design of indicator sets.

Finally, bottom-up approaches have the critical function of involving stakeholders and communities in the process of developing this and all parts of the decision-making process. A participative approach will not always lead to the formulation of clear criteria but may instead lead to the discussion of goals, objectives, and important local issues which can lead either to the direct creation of indicators or the qualities that are important for criteria. In fact, in bottom-up approaches the criteria may often be of a more implicit nature. This information can be translated into criteria by specialists involved in the process. In any case, the design and identification of a framework and the objectives of the measurement process typically need local validation (Reed et al., 2005; Fraser et al., 2006) and stakeholder participation can determine much of the conceptual structure for the design of indicator sets.

Important research is needed into the combination of these methods and as framework models become increasingly consistent, a range of multifaceted methods will emerge. Once more, it is important to state that combinatorial research is beyond the scope of this paper which focuses principally on the phase of comparison and consolidation. Consultation of experts is done obliquely through the quantification of literature references. Finally, while a framework is not presented a simple two-part (scientific or end-user), pre-framework structure is used here only as form of presenting the results of the synthesis of the criteria.

4. Methodology

Within the discussion of criteria, this article deals principally with the use of classification methodology in order to synthesize and propose some level of standardization for criteria within the sustainability/water resource use nexus. With a review of the literature, it was possible to indicate some tendencies in support of certain criteria as possible candidates for status as "core criteria." These are criteria that have appeared in a vast array of studies and methods as well as reflect the results of studies from other sectors and conform to the Bellagio principles. As with the indicators themselves, the number of criteria in use has become numerous and needs constant revision and consolidation. That is the purpose of this study. The synthesis presented represents an example of a process of assessment and consolidation within the sustainability and water resource use sector and it is not meant to be a comprehensive revision of criteria or indicators from all areas connected to sustainable development. It was observed that consolidation is essential, and this assessment would be a valuable contribution in maintaining the overall quality of criteria in common use.

The process of synthesis started with a bibliographical search in order to acquire a significantly representative selection of criteria focusing on the domains of sustainability and water resource management. It included publications presented by international institutions and national governments, as well as ones addressed by the scientific community in peer reviewed international journals. In total, 74 sources were identified in literature revision and keyword searches on Google Scholar, Web of Science, and Scopus. These were examined and found to contain a total of 346 mentions of criteria used for indicator assessment. These sources include publications from internationally recognized institutions that are renowned for their reliable work with indicators, such as the CBD (1999), EEA (2005), FAO (1999), GRI (2002), IISD (2008), OECD (2003), UN (2007), UNEP (2006), US EPA (2000), US GAO (2004), World Bank (2000), WHO (2002) and WWAP (2006).

This study also examined a significant number of peer reviewed scientific papers (Aveline et al., 2009; Bélanger et al., 2012; Bringhenti et al., 2011; Cloquell-Ballester et al., 2006; Gudmundsson, 2010; Kurka & Blackwood, 2013; Meul et al., 2009, Niemeijer & de Groot 2008). However, the majority of the publications analyzed by this research provided insufficient detail about the criteria selection process, reducing the possibility of scientific replication (Aveline et al., 2009).

Most of the sources contain a list of criteria for the evaluation of indicators related to water resources, fishery and agriculture but other sectors were considered including transportation, forestry, health, energy, biodiversity and planning. Studies that were applied on multiple scales including local/national/international were also considered. Five meta-reviews of criteria to select indicators were also analyzed, namely: WWAP (2006), UNEP (2006), Niemeijer & Groot (2008) and Kurka and Blackwood (2013).

A matrix of the 346 criteria identified in this study was built in order to perform classification analysis in order to synthesize and propose possible standardization. The name and definition of the criteria were transferred from the original sources to the matrix and each criterion was examined in order to demonstrate the levels of overlap, redundancy and ambiguity. This synthetic examination revealed that the 346 criteria were in fact 60 unique criteria. Redundancies were principally found to be either the use of the same name but with different definitions or the use of different names with definitions indicating that they were, in fact, the same criterion. The scale of the reduction from 364 to 60 demonstrates the significant levels of redundancy.

The objective of this study was to clarify how criterion are multiplying and thus diminishing effectiveness due to high levels of redundancy. Proposals for standardization are tentative and limited at this time but an essential first step to creating a multifaceted method for the design of criteria sets for the sector. The study proposes a standard name and a description for each criterion, based on the ones presented by the sources analyzed and the intrinsic aspects of each criterion. Furthermore, quantification based on the number of citations mentioned was realized to support the relevancy and standardization process (Lutz & Hans-Dieter, 2008; Radicchi & Castellano, 2012). The number of sources that consider each criterion in question as relevant was counted and the more times a criterion was cited indicated that they were stronger candidates for the function of “core criteria” to be utilized in future framework development.

In a further effort to facilitate the use of the 60 indicators with their relevance ranking, the criteria were distributed into two pre-framework groups one focused on scientific relevance and accuracy and of more probable interest for top-down methods and an end-user group which incorporates Kurtz et al’s (2001) notions of feasibility and utility and these would typically be associated with bottom-up approaches. This grouping method is similar to that used by WWAP (2006). Both groups present highly cited criteria that indicate a certain level of standardization though this fact is hidden by the inconsistent use of criteria names and definitions.

5. Results and Discussion

In the process of classification analysis, it was first noticed that what this study calls "criteria" is sometimes defined by other publications as "guidelines", "requirements", "indicator quality", or "desirable properties" among other terminologies referring to elements that should be considered for the evaluation of the quality of an indicator or a set of indicators. The comparative analysis of each of these "criteria" performed by this study makes clear that little use of standard nomenclature or definitions exists. The same criterion may be called by different names, and similarly-named criteria may have different definitions (meaning, that they are different criteria). Therefore, the names and the definitions of the criteria proposed here (Annexes 1 and 2) bring some clarity in a field that lacks standardization (Niemeijer & de Groot, 2008; Gudmundsson, 2010). Nevertheless, they are not exempt from certain levels of overlap, redundancy or ambiguity and could be further improved in future studies.

A total of 205 mentions of criteria classified as end-use oriented and 141 mentions of criteria with a more scientific orientation were identified (Annexes 1 and 2). These in fact represent 35 different end-use criteria and 25 scientific-focused criteria once synthesized and consolidated. This points to a broader range of end-use type criteria in the literature and this may represent the fact that local stakeholder's issues and objectives are, in general, diverse in nature and it would be consistent that criteria representing end-use/bottom-up approaches are too. In the development of a framework for criteria selection, it is important to be aware of the balance between these two approaches (Bockstaller and Girardin, 2003; Cloquell-Ballester et al., 2006; Meul et al., 2009). Reed et al., (2005) consider these two approaches the basic starting point for framework design. Indicators must first "accurately and objectively measure" and second "it must be possible for local users to apply them." Following this logic, the criteria synthesis was presented in two groups.

Figure 1 below presents the synthesized criteria divided into the two approaches with full tables presented in Annex 1 & 2. The rankings of probable relevance are identified by the citation number realized in this study. Data availability was considered to be the most relevant criterion overall, mentioned by 31 different sources. It is a crucial criterion, mainly because if data is not available it is likely that the indicators will not be used. It is probable that such a criterion will be a core indicator in most frameworks. The simple fact is that for accurate measurements, indicators require the availability of regular and reliable data sources.

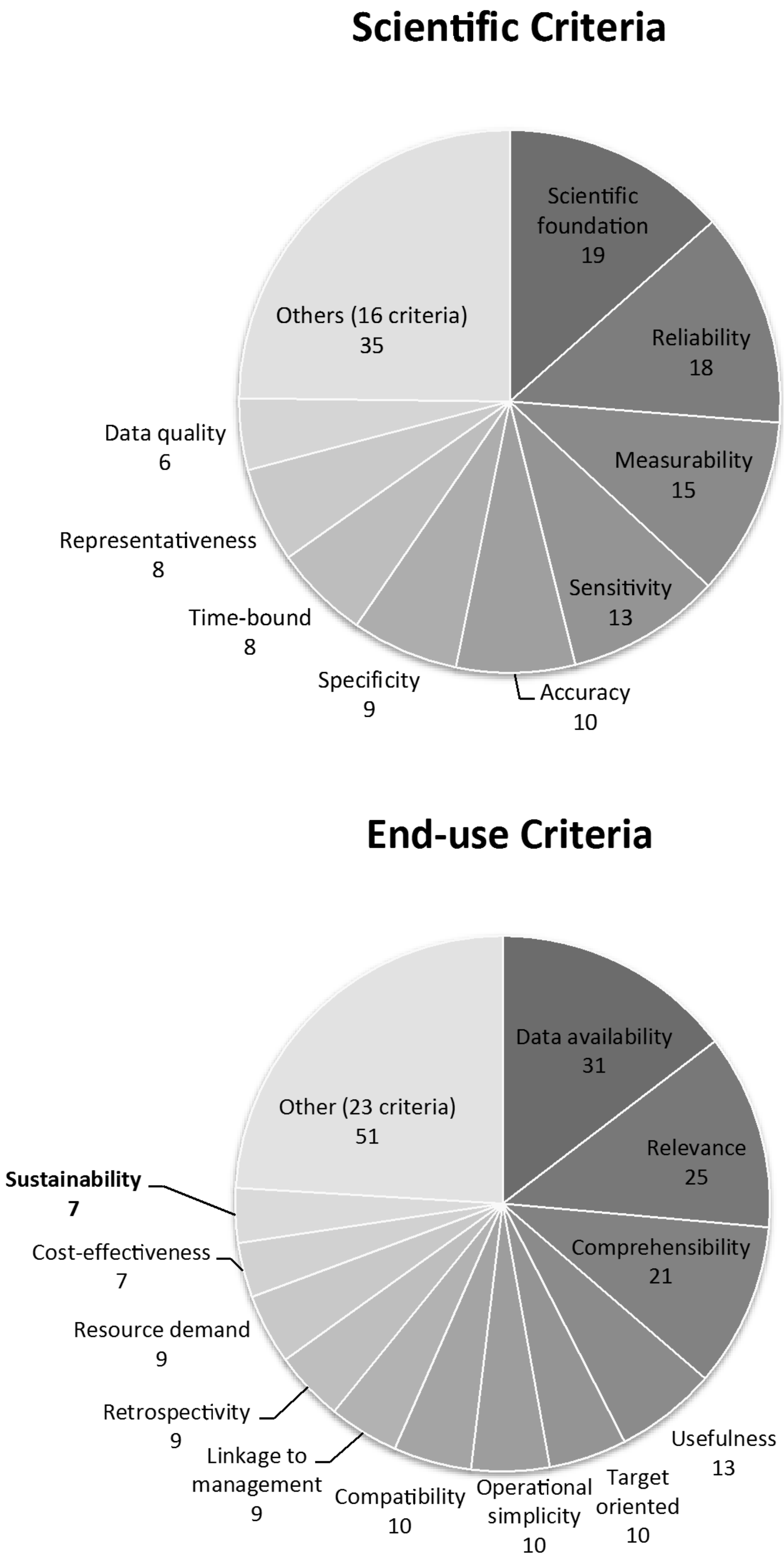

Figure 1 – Most relevant scientific and end-use criteria, indicating the number of mentions.

Scientific Foundation was the second ranked criteria and the highest ranking in scientific approach. Nineteen different sources mentioned it as one of the criteria that should be used to assess indicators. This criterion aims to ensure a solid and concrete scientific relevance to the selection of indicators. Scientific Foundation as a criterion assesses the extent to which an indicator is based on currently sound and internationally accepted theoretical, conceptual, technical, and scientific standards and principles. As with Data Availability, this level of citation indicates a probable core criterion.

Relevance and Comprehensibility are also among the most mentioned criteria. There were 25 mentions of the former and 21 mentions of the latter. They are suggested here as part of the end-user approach. The criterion Relevance corresponds to the extent to which an indicator is related or connected to the matter in hand. The criterion Comprehensibility addresses the extent to which the indicator can be understood by the target audience. Relevance acquires the status of a core quality as its absence clearly defeats the purpose of the measurement and Comprehensibility is the first of a number of criteria that touch on the importance of transforming indicators into effective action and for this reason appears in almost all frameworks such as DSPIR where the "Response" category makes this explicit.

When considering the ranking of the criteria, it is important to consider that the most cited criteria tend to be divergent or universal in application (i.e. applicable to many areas, situations and scales) while less cited may often be convergent on a particular stress or relationship (Cloquell-Ballester et al., 2006). In developing a set of criteria, a search for balance will typically lead to both types of criteria though convergent criteria will depend more on local issues and are therefore less likely to be considered general or core criteria.

Taking together the most cited criteria (10 or more citations) a possible list of core criteria emerges:

- Data availability, 31 citations
- Relevance, 25 citations
- Comprehensibility, 21 citations
- Scientific foundation, 19 citations
- Reliability, 18 citations
- Measurability, 15 citations
- Sensitivity, 13 citations
- Usefulness, 13 citations
- Accuracy, 10 citations
- Target oriented, 10 citations
- Operational simplicity, 10 citations
- Compatibility, 10 citations

These criteria broadly confirm the results of a similar study by National Centre for Health Outcomes Development (NCHOD, 2005), a study realized in a different sector. It also reflects Hák et al.'s list of "Criteria for Methodological Strength of Indicators" (2012: 56-57) and their revision of key criteria used in developing indicators for the Sustainable Development Goals (Hák et al., 2016). This indicates the importance of these criteria but also points to a certain generic character of some of these core criteria. In general, this list indicates candidates for wide use in most criteria frameworks, and if some of them are left out of a framework for being considered too obvious, it is important that these omissions are documented and expressed so that those consulting the indicators understand this. Core criteria, while possibly obvious in nature, are capable of invalidating any measurement process if not taken into consideration.

The number of criteria used to assess indicators is also important, both from a scientific and from an end-user standpoint (Cloquell-Ballester et al., 2006). Using a greater number of criteria can increase the quality of the assessment by assessing validity from a greater number of angles (Niemeijer & de Groot, 2008). However, the adoption of too many criteria could increase the cost and complexity of the selection process. It is therefore common for indicator researchers to mention the need for conciseness.

In summation, once comparative studies have consolidated criteria, it is recommended that the final selection of a criteria set be based on a combination of expert consultations and a collaborative, bottom-up process, guided by a logical framework to guarantee a wholistic and consistent grouping of factors. In turn, this framework should, at minimum, work with a restricted number of criteria that while not necessarily incorporating an equal number of scientific and end-user criteria, should have significant representation from both groups as well as cover most of the core-criteria identified. Developing and exploring how possible combinatorial methods might work with respect to these guidelines, will represent an interesting line of inquiry for future studies.

6. Conclusions

Indicators are key tools that help the society to monitor progress and trends on the path towards the sustainable use and management of natural resources. The application of proper criteria and/or criteria sets to assess indicators is an important part of the process of determining quality and in the design of the indicator set. Nevertheless, the identification and selection of criteria and indicators related to water resource management is currently hampered by a lack of consolidation and standardization.

Our study aimed to demonstrate the need for synthesis and consolidation within the sector and to present an initial proposal for standardization and a tentative ranking of indicators that might be classified as core criteria. In order to do this, the research carried out a representative literature review of 74 sources to evaluate the level of redundancy of commonly used criteria. The findings of the study revealed that the 346 criteria identified in the literature were in fact 60 different criteria. The study initiated a standardization process by indicating names and a description for each criterion, aiming to provide more clarity and reduce ambiguity. The 60 criteria were finally organized and presented in a manner to facilitate their use in criteria frameworks to be developed by combinatorial methods. First, there were divided into two principal approaches (scientific and end-use) and then ranked according to citation to identify possible core criteria. This provides a solid foundation for the design of multifaceted selection frameworks that combine both top-down and bottom-up approaches.

The results can be summed up as a series of guidelines that can aid in the structuring process of indicator sets. First, it is important to eliminate redundancies. Second, it is recommended to utilize a balanced mix of criteria from both top-down and bottom-up approaches. Third, it is critical to be aware that some criteria are core to the basic functionality of indicator use and the transformation of the data acquired into positive actions. Fourth, there is an optimal number of indicators that can be used in a set before it become unwieldly. Lastly, structuring criteria choice with an overall logical framework will help give balance and completeness to the set.

The criteria synthesis presented by this study could be considered a relevant contribution to the development and use of methods for indicator selection since no previous work was found that has conducted such a broad, up-to-date review for this sector. This synthesis was carried out in a transparent and replicable manner, so that it can be advanced with the incorporation of new sources, new criteria and/or regular updates. Furthermore, the tables of criteria, ranked according to their citation number and with core-criteria identified, can be used in future studies that work to develop criteria frameworks in a more consistent and effective manner.

REFERENCES

Aveline, A., Rousseau, M. L., Guichard, L., Laurent, M., & Bockstaller, C. (2009). Evaluating an environmental indicator: Case study of MERLIN, a method for assessing the risk of nitrate leaching. *Agricultural Systems*, 100(1-3), 22–30. doi:10.1016/j.agsy.2008.12.001

Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., & Murphy J. (2002). Guidebook to decision-making methods. USA: Department of Energy.

Becker, J. (2010). Use of backcasting to integrate indicators with principles of sustainability. *International Journal of Sustainable Development & World Ecology*, *17*(3), 189–197.

Bélanger, V., Vanasse, A., Parent, D., Allard, G., & Pellerin, D. (2012). Development of agri-environmental indicators to assess dairy farm sustainability in Quebec, Eastern Canada. *Ecological Indicators*, 23, 421–430. doi:10.1016/j.ecolind.2012.04.027

Bell, S., & Morse, S. (2003). Learning from experience in sustainability. In: Proceedings International Sustainable Development Research Conference 2003, 24-25 Mar 2003, Nottingham, UK.

BNIA - Baltimore Neighborhood Indicators Alliance. (2006). *Vital Signs IV: Measuring Baltimore's progress toward strong neighborhoods and a thriving city* (p. 106). Retrieved from http://cdm16352.contentdm.oclc.org/cdm/ref/collection/p15224coll6/id/1744

Bockstaller, C., & Girardin, P. (2003). How to validate environmental indicators. *Agricultural Systems*, 76(2), 639–653. doi:10.1016/S0308-521X(02)00053-7

Bockstaller, C., Guichard, L., & Makowski, D. (2008). Agri-environmental indicators to assess cropping and farming systems: a review. *Agronomy for Sustainable Development*, *28*(1), 139–149. Retrieved from http://link.springer.com/chapter/10.1007/978-90-481-2666-8_44

Breslow, S. J., Allen, M., Holstein, D., Sojka, B., Barnea, R., Basurto, X., ... & Donatuto, J. (2017). Evaluating indicators of human well-being for ecosystem-based management. Ecosystem Health and Sustainability, 3(12), 1-18.

Bringhenti, J. R., Zandonade, E., & Günther, W. M. R. (2011). Selection and validation of indicators for programs selective collection evaluation with social inclusion. *Resources, Conservation and Recycling*, 55(11), 876–884. doi:10.1016/j.resconrec.2011.04.010

Buchholz ,T., Luzadis, V. A., & Volk, T. A. (2009). Sustainability criteria for bioenergy systems: results from an expert survey. *Journal of Cleaner Production*, 17, S86–98.

Butler, D., Jowitt, P., Ashley, R., Blackwood, D., Davies, J., & Oltean-Dumbrava, C. (2003). SWARD: decision support processes for the UK. *Management of Environmental Quality*, 14, 444–59.

Calliera, M., Marchis, A., Bollmohr, S., Sacchettini, G., Lamastra, L., & Capri, E. (2013). A process to provide harmonised criteria for the selection of indicators for pesticide risk reduction within the framework of the sustainable use directive. Pest management science, 69(4), 451-456.

Castillo, H., & Pitfield, D. E. (2010). ELASTIC – A methodological framework for identifying and selecting sustainable transport indicators. Transportation Research Part D: Transport and Environment, 15(4), 179-188.

CBD – Convention on Biological Diversity. (1999). Development of Indicators of Biological Diversity. Nairobi: Convention on Biological Diversity, subsidiary Body on Scientific, Technical and Technological Advice. Report No. UNEP/CBD/SBSTTA/5/12, 14 pp

Clark, W., & Dickson, N. (1999). The global environmental assessment project: Learning from efforts to link science and policy in an interdependent world. *Acclimations*, 8, 6–7. Retrieved from http://www.hks.harvard.edu/gea/pubs/99art_wc_geaacc.pdf

Cloquell-Ballester, V. A., Monterde-Díaz, R., & Santamarina-Siurana, M. C. (2006). Indicators validation for the improvement of environmental and social impact quantitative assessment. *Environmental Impact Assessment Review*, 26(1), 79–105. doi:10.1016/j.eiar.2005.06.002

Dahl, A.L. (2012). Achievements and gaps in indicators for sustainability. Ecological Indicators 17 (2012) 14–19.

Deming, M.E. & Swaffield, S. (2011). Landscape Architecture Research: Inquiry, Strategy, Design. Hoboken: John Wiley & Sons, Inc.

Doukas, H.C., Andreas, B.M., & Psarras, J. E. (2007). Multi-criteria decision aid for the formulation of sustainable technological energy priorities using linguistic variables. *European Journal of Operational Research*, 182:844–55.

EC - Environmental signals. (2003). Canada's national environmental indicator series 2003. Environment Canada. Retrieved from http://www.ec.gc.ca/soer-ree/English/Indicator_series/de- fault.cfm#pic

EEA - European Environment Agency. (2003). Europe's environment: The third assessment: Environmental assessment report no. 10. Copenhagen: European Environment Agency. Retieved from http://reports.eea.eu.int/environmental_assessment_report_2003_10/en/tab_content_RLR, on 2 October 2004.

EEA - European Environment Agency. (2005). EEA Core Set of Indicators—Guide. European Environment Agency, Copenhagen. Report No. 1/2005, 37 pp.

FAO - Food and Agriculture organization of the United Nations. (1999). Indicators for Sustainable Development of Marine Capture Fisheries. Rome: Food and Agriculture Organigation (FAO). Retrieved from ftp://ftp.fao.org/docrep/fao/004/x3307e/x3307e00.pdf

Fraser, E. D. G., Dougill, A. J., Mabee, W. E., Reed, M., & McAlpine, P. (2006). Bottom up and top down: analysis of participatory processes for sustainability indicator identifi- cation as a pathway to community empowerment and sustainable environmental management. *Journal of Environmental Management*, 78:114–27.

Graymore, M. L. M., Sipe, N. G., & Rickson, R. E. (2008). Regional sustainability: How useful are current tools of sustainability assessment at the regional scale? *Ecological Economics*, *67*(3), 362-372. doi:10.1016/j.ecolecon.2008.06.002

Graymore, M. L. M.,Wallis, A. M., & Richards, A. J. (2009). An index of regional sustainability: a GIS-based multiple criteria analysis decision support system for progressing sustainability. *Ecological Complexity*, 6, 453–62.

GRI - Global Reporting Initiative. (2002). Sustainability report- ing guidelines 2002. Retrieved from http://www. globalreporting.org/guidelines/2002/contents.asp, on 7 April 2004

Gudmundsson, H. (2010). *Criteria and methods for indicator assessment and validation - a review of general and sustainable transport related indicator criteria and how to apply them*. Retrieved from http://cost356.inrets.fr/pub/reference/reports/C356_2.2_report_criteria_HG_220410.pdf

Hák, T., Moldan, B., & Dahl, A. L. (Eds.). (2012). Sustainability indicators: a scientific assessment (Vol. 67). Island Press.

Hák, T., Janoušková, S., & Moldan, B. (2016). Sustainable Development Goals: A need for relevant indicators. Ecological Indicators, 60, 565-573.

Hardi, P., & Terrence Z. (1997). eds. Assessing sustainable development: principles in practice. Winnipeg: International Institute for Sustainable Development, Winnipeg, Saskatchewan. Retrieved from https://www.iisd.org/pdf/bellagio.pdf

Heink, U., & Kowarik, I. (2010). What are indicators? On the definition of indicators in ecology and environmental planning. *Ecological Indicators*, 10(3), 584–593. http://doi.org/10.1016/j.ecolind.2009.09.009

IISD - International Institute for Sustainable Developed. (2008). Bellagio STAMP: Sustainability Assessment and Measurement Principles. Winnipeg: IISD. Retrieved from http://www.iisd.org/pdf/2009/brochure_bellagiostamp.pdf

ITFM - Intergovernmental Task Force on Monitoring Water Quality. (1995) Water-quality monitoring in the United States. U.S. Geological Survey, Water Information Coordination Program, Washington. http://acwi.gov/appendixes/index.html

James, C. A., Kershner, J., O'Neill, S., & Levin, P. S. (2012). A methodology for evaluating and ranking water quantity indicators in support of ecosystem-based management. *Environmental Management*, *49*(3), 703–19. doi:10.1007/s00267-012-9808-7

Juwana, I., Muttil, N., & Perera, B. J. C. (2012). Indicator-based water sustainability assessment - a review. *The Science of the Total Environment*, 438, 357–71. doi:10.1016/j.scitotenv.2012.08.093

Khadka, C., & Vacik, H. (2012). Comparing a top-down and bottom-up approach in the identification of criteria and indicators for sustainable community forest management in Nepal., Forestry: An International Journal of Forest Research, 85(1), 145-158.

Klug, H., & Kmoch, A. (2014). Operationalizing environmental indicators for real time multi-purpose decision making and action support. *Ecological Modelling*, *295*, 66–74. doi:10.1016/j.ecolmodel.2014.04.009

Kurka, T., & Blackwood, D. (2013). Participatory selection of sustainability criteria and indicators for bioenergy developments. *Renewable and Sustainable Energy Reviews*, 24, 92–102. doi:10.1016/j.rser.2013.03.062

Kurtz, J.C., Jackson, L.E., & Fisher, W.S. (2001). Strategies for Evaluating Indicators Based on Guidelines from the Environmental Protection Agency's Office of Research and Development. *Ecological Indicators*, 1(1), 49–60.

Lattimore, B., Smith, C. T., Titus, B. D., Stupak, I., & Egnell, G. (2009). Environmental factors in woodfuel production: opportunities, risks, and criteria and indicators for sustainable practices. *Biomass & Bioenergy*, 33, 1321–42.

Lutz, B., & Hans-Dieter, D. (2008). What do citation counts measure? A review of studies on citing behavior, *Journal of Documentation*, 64(1),45-80.

Lyytimäki, J., Tapio, P., Varho, V., & Söderman, T. (2013). The use, non-use and misuse of indicators in sustainability assessment and communication. *International Journal of Sustainable Development & World Ecology*, *20*(5), 385–393.

Mendoza, G. A., & Prabhu, R. (2000). Development of a methodology for selecting criteria and indicators of sustainable forest management: a case study on participatory assessment. Environmental management, 26(6), 659-673.

Mascarenhas, A., Nunes, L. M., & Ramos, T. B. (2015). Selection of sustainability indicators for planning combining stakeholders' participation and data reduction techniques. Journal of Cleaner Production, 92, 295-307.

McCool, S. F., & Stankey, G. H. (2004). Indicators of sustainability: Challenges and opportunities at the interface of science and policy. *Environmental Management*, *33*(3), 294–305. doi:10.1007/s00267-003-0084-4

Meul, M., Nevens, F., & Reheul, D. (2009). Validating sustainability indicators: Focus on ecological aspects of Flemish dairy farms. *Ecological Indicators*, 9(2), 284–295. doi:10.1016/j.ecolind.2008.05.007

Mitchell, G., May, A., & McDonald, A. (1995). PICABUE: a methodological framework for the development of indicators of sustainable development. The International Journal of Sustainable Development & World Ecology, 2(2), 104-123.

Nicholson, E., Collen, B., Barausse, A., Blanchard, J. L., Costelloe, B. T., Sullivan, K. M. E., … Milner-Gulland, E. J. (2012). Making robust policy decisions using global biodiversity indicators. *PLoS ONE*, 7(7). doi:10.1371/journal.,pone.0041128

Niemeijer, D., & de Groot, R. S. (2008). A conceptual framework for selecting environmental indicator sets. *Ecological Indicators*, 8(1), 14–25. doi:10.1016/j.ecolind.2006.11.012

NCHOD 2005. Compendium of Clinical and Health Indicators User Guide. National Centre for Health Outcomes Development (NCHOD), London site - London School of Hygiene and Tropical Medicine.

NRC - National Research Council. (2000). Ecological Indicators for the Nation. Washington, DC: National Academy Press. Retrieved from http://books.nap.edu/books/0309068452/html/1.html.

OECD - Organization for Economic Co-operation and Development. (2001). Environmental Indicators: Towards Sustainable Development. Paris: OECD, 155 pp.

OECD - Organization for Economic Co-operation and Development. (2003). OECD environmental indicators: Development, measurement and use, Reference paper. Paris: OECD. Retrieved form http://www. oecd.org/dataoecd/7/47/24993546.pdf.

OECD - Organization for Economic Cooperation and Development. (1994). *Environmental indicators: OECD core set*. Paris

Olsthoorn, X., Tyteca, D., Wehrmeyer, W., & Wagner, M. (2001). Environmental indicators for business: a review of the literature and standardisation methods. *Journal of Cleaner Production*, 9, 453–63.

Parris, T. M., & Kates, R. W. (2003). Characterizing and Measuring Sustainable Development. *Annual Review of Environment and Resources*, 28(1), 559–586. doi:10.1146/annurev.energy.28.050302.105551

Pastille Consortium (The). (2002). Indicators into action: Local sustainability indicator sets in their context. Final report. London: London School of Economics http://www.lse. ac.uk/Depts/geography/Pastille/FinalReportWeb.pdf.

Prescott-Allen, R. (2001). The wellbeing of nations. Washington DC: Island Press.

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics*, *6*(1), 121–130. doi:10.1016/j.joi.2011.09.002

Reed, M., Fraser, E. D., Morse, S., & Dougill, A. J. (2005). Integrating methods for developing sustainability indicators to facilitate learning and action. Ecology and society, 10(1).

Rovere, E. L. L., Soares, J. B., Oliveira, L. B., & Lauria, T. (2010). Sustainable expansion of electricity sector: sustainability indicators as an instrument to support decision making. *Renewable & Sustainable Energy Reviews*, 14, 422–9.

Segnestam, L. (2002). Indicators of environment and sustainable development: theories and practical experience. Washington: World Bank. Retrieved from http://siteresources.worldbank.org/INTEEI/936217-1115801208804/20486265/IndicatorsofEnvironmentandSustainableDevelopment2003

Shmelev, S. E., & Rodríguez-Labajos, B. (2009). Dynamic multidimensional assessment of sustainability at the macro level: the case of Austria. *Ecological Economics*, 68, 2560–73.

Singh, A., Bedrich M., & Thomas L. (2002). Making science for sustainable development more policy relevant: New tools for analysis. ICSU Series on Science for Sustainable Development, no. 8. International Council for Science Retrieved from http://grid2.cr.usgs.gov/publications/newtools.pdf.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, *9*(2), 189-212. doi:10.1016/j.ecolind.2008.05.011

SNZ - Statistics New Zealand. (2002) Socio-economic indicators for the environment. Environmental Statistics Team, Christchurch. http://www2.stats.govt.nz/domino/external/web/prod_serv.nsf/092edeb76ed5aa6bcc256afe0081d84e/94b7f3198c9d9111cc256c1500171ea6?OpenDocument

Spangenberg, J. H. (2008). Second order governance: Learning processes to identify indicators. *Corporate Social Responsibility and Environmental Management*, *15*(3), 125–139. http://doi.org/10.1002/csr.137

Spreng, D. & Wils, A. (2000) Indicators of Sustainability: Indicators in Various Scientific Disciplines, *Alliance for Global Sustainability*. AGS Report.

Tanguay, G. A., Rajaonson, J., & Therrien, M. C. (2013). Sustainable tourism indicators: Selection criteria for policy implementation and scientific recognition. Journal of Sustainable Tourism, 21(6), 862-879.

UN - United Nations. (2007). *Indicators of Sustainable Development : Guidelines and Methodologies*. New York: United Nations. Retrieved from http://www.un.org/esa/sustdev/natlinfo/indicators/guidelines.pdf

UNEP - United Nations Environment Programme. (2006). *Environmental Indicators for North America*. Nairobi: UNEP. Retrieved from http://www.unep.org/pdf/NA_Indicators_FullVersion.pdf

US EPA – United States Environmental Protection Agency. (2000). *Evaluation Guidelines For Ecological Indicators*. Washington. Retrieved from http://www.epa.gov/emap/html/pubs/docs/resdocs/ecol_ind.pdf

US GAO - United States Government Accountability Office. (2004). Environmental indicators: Better coordination is needed to develop environmental indica- tors sets that inform decisions. Report to Congressional requesters. US Government Accountability Office.

Vera, I., & Langlois, L. (2007). Energy indicators for sustainable development. *Energy*, 32, 875–82.

Wang, J., Jing, Y., Zhang, C., & Zhao, J. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable & Sustainable Energy Reviews*, 13,2263–78.

WHO - World Health Organization. (2002). *Health in Sustainable Development Planning: The Role of Indicators* (p. 42). Geneva.

World Bank. (2000). *Developing Indicators: Lessons Learned from Central America* (p. 22). Washington. Retrieved from http://siteresources.worldbank.org/INTEEI/811099-1115809852605/20486445/DevelopingIndicatorsLessonsLearnedFromCentralAmerica2001part1.pdf

WWAP - World Water Assessment Programme. (2003). The United Nations World Water Development Report: Water for People Water for life. Paris: UNESCO, and London: Earthscan.

WWAP - World Water Assessment Programme. (2006). The United Nations World Water Development Report 2: Water - A Shared Responsibility. Paris: UNESCO, and London: Earthscan.