D2.2 Report on standardized metrics/indicators for monitoring the efficiency of ES/NC based measures

Stefan Schmidt, Anne-Christine Mupepele, Heera Lee, James Paterson, Ralf Seppelt, Sven Lautenbach, Carsten F. Dormann, Martin Volk
UFZ, ALU, UBO, University of Edinburgh

10 December 2014

Ecosystem Science for Policy & Practice

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 308393.

Prepared under contract from the European Commission

Project reference: 308393

Collaborative project

FP7 Environment

Project acronym: **OPERAs**

Project full title: **Operational Potential of Ecosystem Research Applications**

Start of the project: 1 December 2012

Duration: 54 months

Project coordinator: The University of Edinburgh

Project website: operas-project.eu

Deliverable title: Report on standardized metrics/indicators for monitoring the efficiency of ES/NC based measures

Deliverable number: 2.2

Nature of the deliverable: Report

Work package responsible: WP2

Partner responsible: Stefan Schmidt

Other partners involved: Stefan Schmidt, Anne-Christine Mupepele, James Paterson, Heera Lee, Ralf Seppelt, Sven Lautenbach, Carsten F. Dormann, Martin Volk, Meriwether Wilson, Ariane Walz, Genevieve Patenaude

Due date of deliverable: Month 24

Actual submission date: Month 25

Deliverable status:

Version	Status	Date	Authors
1.0	Draft	10 December 2014	Stefan Schmidt, UFZ

Contents

List of Figures

1 Introduction

The ecosystem service (ES) framework analyses goods and services obtained by nature to support human well-being. It is increasingly used in publications not primarily interested in the concept itself, suggesting that it is becoming an increasingly mainstream and accepted idea within the broader scientific community (Abson et al. 2014). Moreover, the ES concept promises a powerful contribution to decision making (Kandziora et al. 2013, Daily et al. 2008, Armsworth et al., 2007). However, challenges remain as the ES concept moves from being primarily a heuristic model towards becoming an explicit management tool. Various guidelines have been developed in order to better integrate ES approaches into decision making, management actions and planning. Most guidelines focused on how to perform assessments (Ash et al. 2010; Ranganathan et al. 2008). Some of those addressing specific user groups, such as economists (Cranford et al. 2011; Hanson et al. 2012), development planners (Kosmus et al. 2012) and conservation scientists (Kettunen et al. 2009; Stolton et al. 2008). These guidelines could contribute to integrate ES approaches into decision making, but they are usually too specific to monitor and compare ES assessments[1]. Furthermore, criteria to measure the progress of ES based measures are neglected in all of the guidelines. To assess the progress of ES research comparison criteria, such as *efficiency* are required. Temporal trends within ES case studies show that *efficiency* becomes more and more a normative character (Abson et al. 2014), i.e. a judgment criteria whether a shift can be deemed as “good”.

The term *efficiency* is gaining widespread recognition in research (Olschewski et al. 2011, Nakhooda et al. 2013, Sprinz 2000), private sector (Oxfam 2012, Hanson et al. 2012) and politics (EU 2013, 2014), although considerable confusion exists on how to define and measure efficiency of ES studies in a standardized way (Gibson et al. 2000, Laurans et al. 2013, Zscheischler et al. 2014). *Efficiency* in its most generic definition determines *the ratio between results achieved (outputs) and resources used (inputs), thus, defines efforts and the endeavor optimization*. Before considering efficiency a requisite is the capacity to produce a desired result. In other words *the quality – or accuracy and completeness – with which an ES study achieved an objective*, i.e. *effectiveness* needs to be proofed. So, for instance the conservation approach of getting the most species per unit area of land protected via a prioritized hotspot scheme have been used in the U.S. as an efficient way to preserve species by focusing on just a few small clusters of critical areas (Kareiva et al. 2003). This strategy, however, would fail to be effective regarding to the protection of species that require vast territories of relatively undisturbed habitats, and thus cannot be contained in hotspots.

Moreover, the determination of what is effective and efficient is strongly context specific and is investigated by a manifold set of indicators in the literature (Sprinz 2000, Phillips et al. 2009, Martin-Lopez et al. 2013, Ruckelshaus et al. 2013, Laurens et al. 2013). In Sprinz et al. (2000), for instance, effectiveness of international environmental regimes in terms of environmental problems are examined along the dimension of use of policy instruments (environmental threshold regulations), represented by the absence of exceeding critical loads in the case of trans-boundary

[1] For more details look at the Milestone 2.1

acidification. In contrast Bagstad et al. (2013) compared ES tools on the basis of five criteria, such as time efforts needed to complete an ES assessment. Both effectiveness and efficiency strongly depending on the objective, i.e. only after the definition of the target it can be assessed what is effective and efficient. Objectives in ES research can be very divers, e.g. validation and comparison of new cost and time saving ES tools (Villa et al. 2014, Bagstad et al. 2013) or policy advice for land use interventions and optimization of ES provisioning (van Wilgen et al. 1998, 2008). Consequently, modularization and standardized structuring forms the foundation for generalizable recommendations on effectiveness and efficiency (Milestone 2.7).

The success of the ES concept as a management tool for human-nature-systems depends on the maturity of operationalizes the concept in practice. To test the degree to which progress have been made a standardized evaluation system and common reporting tool is required. Therefore, indicators for effectiveness and efficiency need to be identified and integrated in a systematic reporting protocol (Task 2.3). This is the foundation for study-crossing gain in knowledge and optimization of the ES concept.

For this report, *Deliverable 2.2*, Task 2.1 (Meta-Analysis) of the OPERAs project, we examined effectiveness and efficiency indicator of ES based measures. Therefore, we first developed a hierarchical framework for a systematic analysis of effectiveness and efficiency indicators. Second, we reviewed major ES databases dealing with globally distributed ES studies respectively projects and identified key indicators that need to be considered for an evaluation of effectiveness and efficiency. Third, the challenge of an evidence-driven approach to estimate the quality of implementation in studies dealing with ES is highlighted. Fourth, we discussed shortcomings and their solutions of identified effectiveness and efficiency indicators as standardized metrics for monitoring. Finally, we show how already existing reporting guidelines – such as the OPERAs *blueprint protocol* – should be extended to enhance monitoring of effectiveness and efficiency as well as to aid informing on evidence.

2 Identification of effectiveness and efficiency indicator

2.1 Framework for analysis

For the identification of effectiveness and efficiency indicators we developed a hierarchical analysis framework of four overarching comparison criteria, see Fig. 1. The hierarchical framework enables to first define the most important dimensions and then more complex, associated sub-categories. It ensures that indicators are extracted by the topic which matters most, not simply the ones that are easiest to measure. Applying a validation of this top-down approach also facilitates to identify areas where data are incomplete, providing a foundation for prioritization of neglected indicator. Within the hierarchical framework the comparison criteria for effectiveness are the (i) *accuracy* and (ii) *completeness* with which an ES study or project *achieved an objective*. For efficiency we used the effort of an ES study or project measured by the (iii) *resources used* to (iv) *achieve an objective*.

Accuracy and *completeness* with which a desired result is achieved refers to the quality of study or project. With the estimation of the *accuracy* the errors of outcomes are valued (ISO 5725). Beside the traditional measures of uncertainty, which are conditional to the method setting, statements about the potential relevance of the model to the context under analysis are important to consider (Spiegelhalter et al. 2013, Wynne 1992). Traditional measures of uncertainty try to capture the uncertainty caused by incomplete system understanding (which processes to include, which processes interact), from imprecise, finite and often sparse data and measurements, and from uncertainty in the baseline inputs and conditions for model runs, including predicted inputs (Jakeman et al. 2006, Walker et al. 2003). Uncertainty measures with respect to the potential relevance express fundamental doubts about the capacity to measure the issue in question (Keynes 1937, Rumsfeld 2002). These measurement inadequacies require to estimate the confidence of the basic understanding or the evidence underlying the study. Mupepele et al. (2014) showed how the evidence base of the effectiveness of ES studies and projects can be estimated. The approach was developed in OPERAs Task 2.1.2 and is explained in more detail in the section: *Evidence-based approach for ES*.

The planed objectives and the actual impact vary considerably between as well as within ES studies. The *completeness* of the target achievement captures the degree to which an effect was accomplished or how far interventions were carried out within a project. To assess the extent of *completeness* the ratio of *achieved objective* to planed target needs to be determined. The *achieved objective* represents the actual outcome and impact of the ES study or project. It refers to the alteration in science, uptake in decision making processes or environmental changes from a study or project (Laurans et al. 2013, Carvill et al. 2012, Oxfam 2012), e.g. innovative methods invented, changes in legislation issues or respectively trees planted. To avoid incommensurability of effectiveness and efficiency between ES studies in our framework different objectives are considered. Furthermore, we included background information on local conditions in the investigation area such as site descriptions that contain drivers and pressures. Thereby we aimed at respecting local peculiarities that may lead to implementations of study/project findings and at improving comparability between studies and projects.

Additionally, *resources used* is an important comparison criterion for efficiency that involves the extent of time and costs used. Costs are not limited to monetary units but involve all other required means. Temporal requirements to achieve the objectives are for instance the study duration or the periodicity of the assessment, if it is repeated (Whitlock et al. 2009). Monetary costs can capture the use of data gathering, involvement of experts, application of tools and methods to reach the objectives based on market prices such as funding and interaction costs to implement the study (Nakhooda et al. 2013). Further means required were subdivided into methods and people involved. Methods represent the design and complexity of the study with a focus on which and how tools, activities and data sources are used (Martin-Lopez et al. 2013, Bagstad et al. 2013), e.g. during a soil sample campaign for carbon measurement to estimate carbon sequestration. People involved encompass the expertise, the number of persons and stakeholder engaged in the study or project (Whitlock et al. 2009). This could be implemented for example by the number of people considered for the assessment differentiated by their background or by measuring the capacity building needs identified during the assessment.

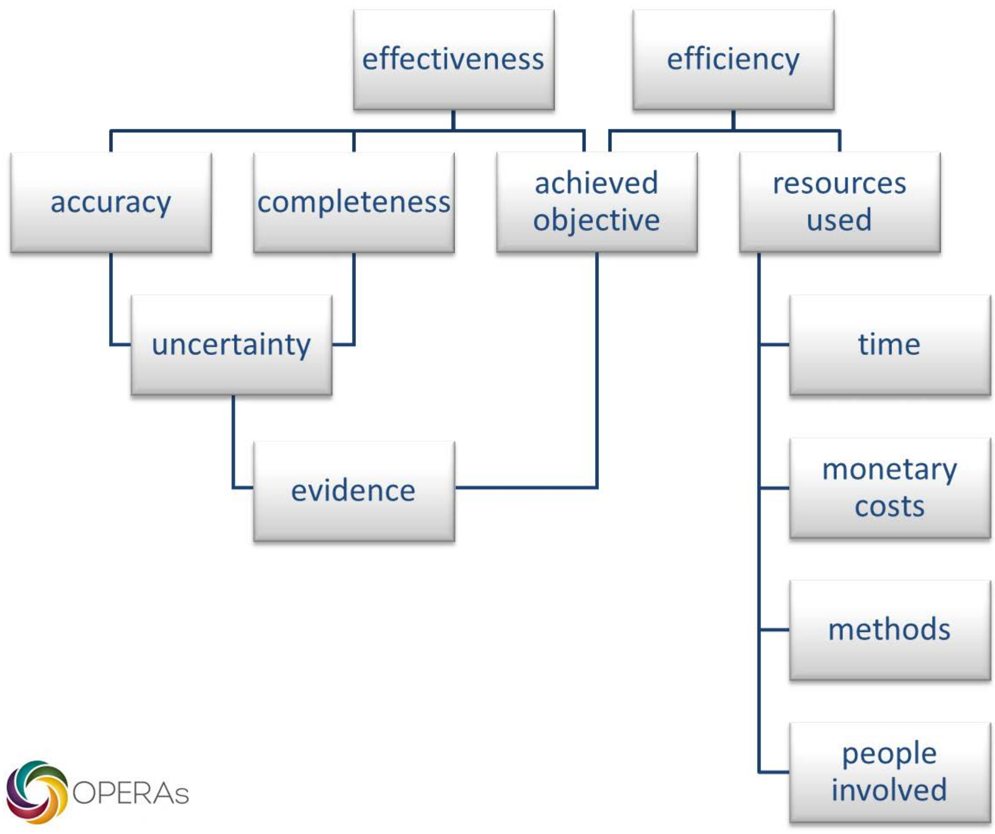

Fig. 1. Comparison criteria for the analysis of effectiveness and efficiency indicator. Notice that efficiency criteria may also be the issue in question for effectiveness analysis, for instance if a study aims to develop innovative cost and time saving ES tools.

2.2 Available Databases

For the identification of effectiveness and efficiency criteria we reviewed major databases deal with globally distributed ES case studies or projects and analyzed them with the above mentioned framework. Databases are a suitable starting point for the investigation of effectiveness and efficiency criteria, as they aim to collect datasets, condense them and improve the information supply on comparable indicators. We only integrated databases that were online available without a narrow regional focus (see Appendix 1). This ensured a comprehensive overview about the complexity of different natural and social systems, avoided biases due to local peculiarities and increased the relevance for a broader audience. To analyse the databases we fist compared their entities which capture effectiveness and efficiency criteria, second examined the respective entries, reflecting study or project attributes.

2.3 Results

Our review shows that the databases analysed provide only limited insights with respect to effectiveness and efficiency of the ES studies and projects. This might be explained by the fact that the purposes of the various databases were highly heterogeneous and none were created with the intention of appraising effectiveness or efficiency specifically. We also observed that the indicators for effectiveness and efficiency shifted considerably with the focus of the databases. While databases which aimed at verify the credibility of ES quantifications (ZEN) or estimate monetary values of ES (e.g. ESVD, MESP, ENVALUE, PES) predominantly feature methodological indicators, the databases that directly seek to establish regulation schemes (IIED) or summarize lessons learned from ongoing assessment processes (CA, TEEB) also show indicators on costs, time and impact. Nevertheless, the databases we examined provided 541 entities that were potentially suitable for a standardized reporting on effectiveness and efficiency. We aggregated these into 30 indicators (Fig. 2).

It is striking that most of the indicators were means to report on efficiency (60% of indicator entries). With 55%, the largest proportion was represented by methodological aspects of the ES studies. Indicators that imply background information of local conditions and drivers occurred in 25%, but they contributed only slightly to cluster the ES studies and projects into comparable subdivisions due to their high thematic variety. Effectiveness indicators for the estimations of outcomes and impacts were rare with only 9%. The database from Goldman et al. (2008), for instance, featured most indicators on outcomes and impact with 47 criteria, but less than the half of the data had entries for more than 5 indicators. Indicators on expertise, capacity building and people involved in the study or project appeared in 7%, followed by indicators on uncertainties (2%) and indicators on statements on objectives (2%). Expectedly, the number of indicators on costs and time issues was very low (each of 1%), representing a challenge for mostly efficiency appraisals.

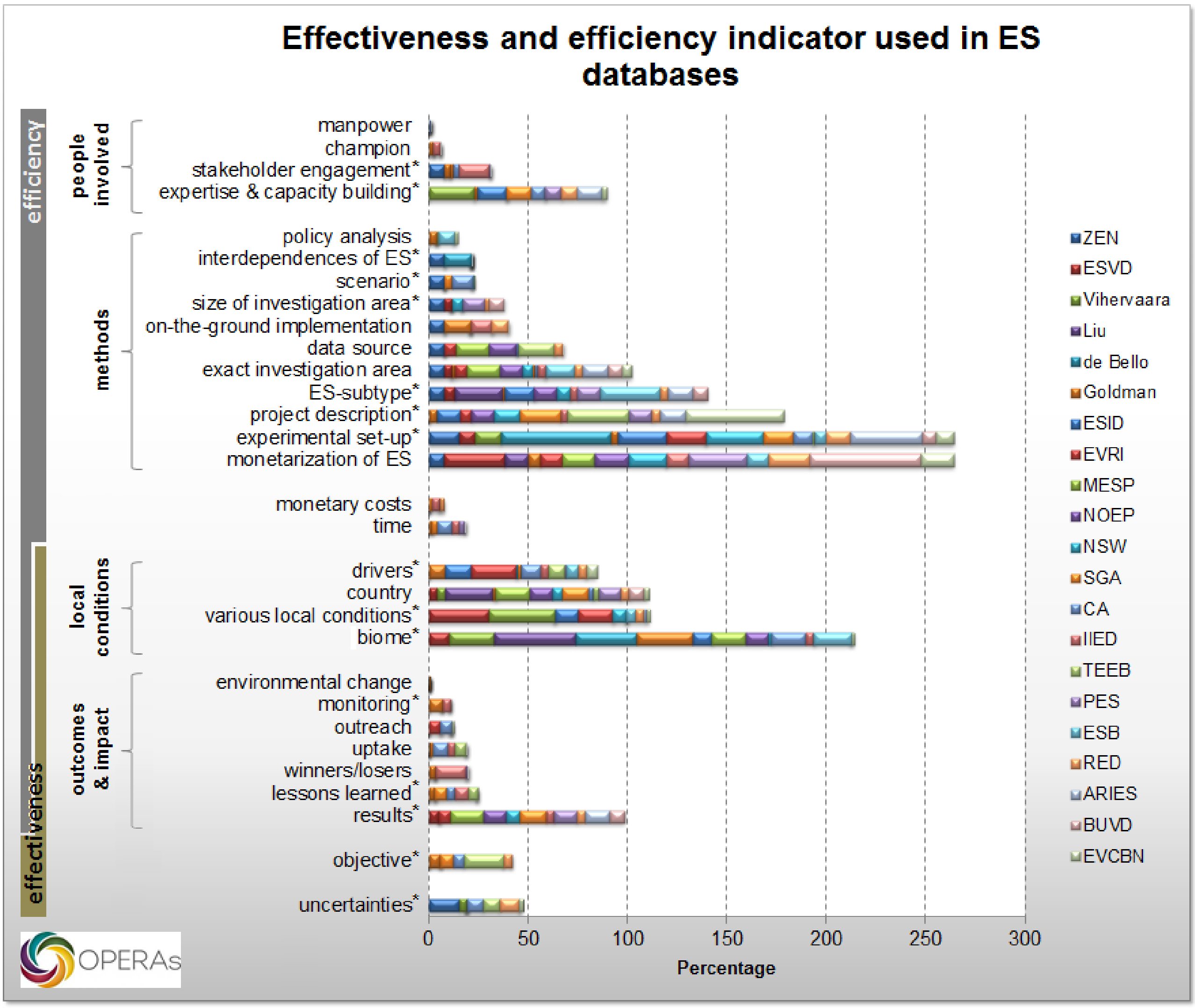

Fig. 2. Overview of 30 effectiveness and efficiency indicators based on 21 global ES databases. The percentage value on the x-axis shows the summed value from all databases entries. On the left side of the figure it is shown how the indicators can be aggregated in terms of effectiveness and efficiency. Additionally, the stars marking indicators that are already included in the OPERAs blueprint protocol.

The numerous amounts of indicators for methods can be attributed to the high diversity of different types and topics, ranging from information on data input over set-up configurations to specific recommendations given for biophysical and socio-economic approaches. Mostly indicators for economic valuation were represented. This supports the increasing popularity of monetary valuation in ES research (TEEB[2], WAVES[3], SEEA[4], de Groot et al. 2012) and the long, continuing discussions about economic valuation as the key tool for a more effective mainstreaming of biodiversity and ES (COP 2010, Heal 2000). Second most frequently observed were indicators on the experimental set-up, an aggregation of entities that consists of input data requirements,

[2] TEEB: The Economics of Ecosystems and Biodiversity http://www.teebweb.org/

[3] WAVES: Wealth Accounting and the Valuation of Ecosystem Services http://www.wavespartnership.org/en

[4] SEEA: System of Environmental-Economic Accounting https://unstats.un.org/unsd/envaccounting/seea.asp

indicators and metrics used as well as general descriptions of methods and tools characteristics. A substantial lower number of indicators allowed conclusions on specific methods, tools and mechanism that led to on-the-ground activities and described activities applied to achieve final goals. Examples in the databases therefor are specific recommendation on easements or major legal tools for actual regulation respectively policy to cause some kind of change; such as creation of new subsidies. Also information on policy analyses were neglected in the ES databases, for instance whether the projects assessed what was needed politically or legally in order to trigger project actions.

In accordance to Laurans et al. (2013) we conclude that insufficient information is available on how ES case studies had an impact on management processes or decisions. The number of studies or projects that monitored study induced environmental changes was scarce as well. However, a few selected studies indicated that:

i) similarly to IPBES (2013), case studies and projects mandated by governments and/or intergovernmental processes were generally more closely aligned with the needs of decision makers, and thus have a kind of “receiving environment‟ for the findings;
ii) case studies and projects that identified and integrated champions in the examination showed a higher uptake of results. Champions are experts that promotes and stand up for scientific results in policy and law making;
iii) policy analysis as a part of the ES study helps to find important entry points for the linkages of scientific findings with management structures and processes, but doesn’t guarantee the implementation in decision making processes;
iv) the policy and advice generation process in an ES assessment is the most time consuming issue – it takes a lot of time to create user-friendly indicators, metrics and visualizations, develop guidelines tailored to the audience, identify champions, assist in applying indicator and guidelines etc.

Only a low number of efficiency indicators were reported on: the expertise of researcher conducting the study, capacity building efforts during the ES study/project, capacity building as an action taken by the study/project, stakeholder engagement as well as manpower in the case study or project. The proficiency, capability and manpower of a research team strongly affected time and cost issues in a study or project. When capacity-building was integrated into the study/project process it can broaden and enhance participation, as well as lead to development of capacity to perform assessments on an ongoing basis (IPBES 2013). Furthermore, the engagement of stakeholders helped at all stages in an assessment process to ensure the credibility, relevance and legitimacy of a study/project, and increases the extent to which findings are reflected in decision making. Recent studies have indicated that stakeholder values are the key to structured policy making with public involvement (Gregory et al. 2001, Gregory 2000). In real terms, Lorenzoni et al. (2000) found for a case study in East Anglia that indicators that had been designed to meet the practical needs of stakeholders worked best.

The documentation of uncertainties is mostly neglected (13 out of 21 databases) and only two databases contained detailed information on qualitatively and quantitatively validation of results. This underlines the findings from Seppelt et al. (2011) and highlights again the importance to report on validity and robustness of scientific results in the face of uncertainties to ensure credibility and relevance for different user groups.

In only five databases entities were integrated that gave insights into the objective underlying the ES studies or projects. None of those used a standardized classification system, so that only with extensive, additional expenses a consistent foundation could be provided to conduct effectiveness or efficiency assessments.

Data on how well and how much monetary costs and time were spent for an ES study/project were sparse in ES databases. At the moment there is only one database dealing with the implementation of payments for watershed markets (IIED) that provides information on time issues (project maturity) as well as funding and partly on what the money is used for in 69 projects. Monetary cost and time requirements are central criteria to estimate efficiency, but also effectiveness. They limit both the investigation scope and methods (Bagstad et al 2013), they determine boundaries for the consolidation and integration of experts and stakeholder (IPBES 2013), thus, affect the quality of the findings. Based on a comparison of 17 ES tools Bagstad et al. (2013) showed that cost and time requirements to run quantitative ES models remain too high to be used in widespread decision making, in contrast to low-cost screening tools that should be more used for scoping due to the risk of oversimplification of environmental complexities.

2.4 Shortcomings and solutions

The ES concept interlinks natural science with social science and decision making. On the one hand this interdisciplinarity is a strength of the concept that enables the analysis of the big picture of how nature fulfils human life and how we in return impact our ecosystems. On the other hand for the examination various disciplinary perspectives and a plethora of input data, scientific methods and policy instruments are required. Consequently, challenges occur that needs to be considered for the monitoring of effectiveness and efficiency. We revealed the following major challenges within our analysis, and then go on to discuss these in more detail:

- a standardized classification system for the aggregation of similarly objectives is needed,
- awareness on the influence of methods and disciplinary influence on study outcomes,
- the process within ES studies reflected by various cooperation and communication requirements needs to be thoroughly assessed,
- consideration of trade-offs between the generality and comparability of effectiveness and efficiency indicators determined by the level of aggregation of indicator groups,
- organized collection of data is required that condense the information oversupply.

The objective of a study limits the examination scope within effectiveness and efficiency analysis can be conducted (Ruckelshaus et al. 2013, Vogel 2012). So, first of all for a standardized reporting on effectiveness and efficiency a commonly used approach of determined objectives is needed. Daily et al. (2009) developed a framework which utilizes the holistic idea of the ES concept and structured important components in a simplified way to be aware on multidimensional and interdisciplinary requirements for the operationalization of the ES concept into practice. Databases, projects and studies dealing with ES rarely applied the ES concept entirely (UK NEA 2011, van Beukering et al. 2003), rather focusing on certain components, e.g. the quantification of ES (Doherty et al. 2010, Phoenix et al. 2008, Marrs et al. 2007), monetary valuation (Zaho et al. 2008, Ingraham et al. 2008), policy analysis (Nilsson et al. 1995). With the framework provided by Daily et al. (2009) standardized objectives for the whole ES concept can be distinguished and similarly ES studies aggregated to establish a consistent foundation for the comparison of effectiveness and efficiency (see Box 1).

Studies that are effective and efficient with respect to the achievement of a certain Daily et al. (2009) component, not necessarily meet the overarching objective of the ES concept.[5] "Component studies" are often dominated by a specific scientific discipline and prevailing methods. There is evidence that ES assessment techniques do not simply uncover but also construct values (Vatn et al. 1994) and should be considered as value-articulating institutions (Jacobs 1997; Vatn 2005, Brondizio et al. 2010). Furthermore, choices of methodological approaches can bias information according to a particular disciplinary perspective, and potentially limiting of how interests of ES beneficiaries are reflected (Martin-Lopez et al. 2013). The dominance of monetary valuation in scientific literature, for instance, entails a privileged position in environmental decision-making process (Carpenter et al. 2009, Daily et al. 2009, Wegner et al. 2011).

Box 1. Classification system for ES objectives

Daily et al. (2009) developed a framework that considers a number of ES simultaneously over scales appropriate to local-, regional-, and national-level resource-management decisions; and it connects the science of quantifying services with valuation and policy work to devise payment schemes and management actions. This framework can be adopted as classification system to distinguish between different objectives of ES studies and projects. Thereby, the following ten groups are considered: i) decisions made by individuals, communities, corporations and governments (decisions), ii) land management, conservation, restauration interventions (actions and scenarios), iii) quantification of ecosystem structures and processes, iv) quantification of ES states and flows, threats, structures or processes, v) development of improved biophysical methods, vi) development of improved economic, cultural models vii) measuring the values of ES to people, viii) altering cultural norms and behaviour, ix) analysis and design of institutions that will guide resources, management and policy, x) integration of conservation in decision making.

Martin-Lopez et al. (2013) argue that ES research requires a variety of methods in order to capture complexities and value pluralities that exist in the systems of interest. The complexity of the ES concept requires a dialogue between natural and social scientists, as well as between academics and policy-makers. This raises another challenge for ES studies, if different components of the Daily et al. (2009) framework from the supply over the demand-sides to the decision making should be considered. The implementation of an ES endeavor is not only choosing a method and

[5] The ultimate goal of the ES concept is to guide the usage of land, water and other elements of natural capital by better understanding and valuing nature supply for human well-being (Daily et al. 2009).

measuring the issue in question; it is also a process with various communication and cooperation requirements. These factors are often not reported in the studies and projects, so cannot be considered in most effectiveness and efficiency analyses, although they might be decisive for achieving the objectives. In Zscheischler et al. (2014) several key factors for success are discussed, for example actor constellation and interaction as well as power balance in research projects. They argue *inter alia* that powerful actors tend to impose their individual interests into the research process with a negative effect on long-term sustain actions.

A further challenge that determines the comparability of effectiveness and efficiency is the aggregation of indicators. Aggregation is the process of combining individual indicator at a given scale, based on the indicators on a more detailed scale (Piñeros- Garcet et al. 2010). Aggregating indicators on a high level provide a basis for comparing larger samples than at a low level, but this presuppose that indicators are combined can be summarized to a single rod of measure (Martínez-Alier 2002). Some of the indicators identified in this analysis are highly aggregated, i.e. they encompass many entities dealing with conflicting disciplinary languages or reflecting different stages in the ES analysis. The efficiency indicator *time*, for example, includes time requirements for conducting the field measurements of the study and time needed to apply study-induced environmental change actions. With our analysis we showed overarching categories as well as examples of more disaggregated indicators that can be used as standardized metrics to monitor effectiveness and efficiency, but according to the complexity of objectives in ES research iterative extension is inevitable. The hierarchical framework of our analysis easily allows it to add further indicators. Therefore, it is important to find an appropriate balance between the levels of indicator aggregation and comparability to ensure valid estimations with the fewest possible number of metrics.

Databases are a suitable starting point to find effectiveness and efficiency criteria. One of the advantages is that they show common summary statistics of collected primary studies in a standardized form to aid reporting. Furthermore, findings from IPBES (2012) emphasize that database provide meta-information relevant for effectiveness and efficiency that is sometimes not available even in the studies. The disadvantage, however, is that databases are context specific and condensing information accordingly to a particularly stressed topic. Thus, information that is neglected in the databases is not necessarily missing in the studies or projects themselves and needs to be cross-checked by a detailed literature review. Although we cannot in-depth re-analyze all 12.000 studies of the synthesized databases, we found evidence throughout the current literature on meta-analyses that there is a paucity of publications reporting on effectiveness and efficiency indicators in the same way we identified (Task 2.1.5, Milestone 2.3, and Zscheischler et al. 2014, Laurans et al. 2013).

3 Evidence-based approach for ES

Decision-makers need to know the evidence base of their actions. It is essential to assess the reliability of our knowledge on effectiveness/efficiency of the proposed intervention. The evidence assesses this reliability and tells us whether our predictions are based on strong evidence, hence very reliable, or very weak evidence, which means that we assume it can work, without knowing it well.

In environmental sciences, the evidence-based concept is well developed in conservation, even though it is still limited in conservation practice. Indeed, there is not at present any guideline for ES science. To apply the evidence-based concept in practice we need a scale to rank study designs commonly used in ES science as well as an assessment of the actual quality of implementation in a specific study. Within OPERAs task 2.1.2 we defined a framework for evidence-based ES science, together with a design scale and a critical-appraisal checklist. Both can be found together with detailed examples on the application by Mupepele et al. (2014).

Evidence-based practice starts with a question or a purpose (Fig. 3). This question is answered in research studies that follow a resolute study design. The study design is the set-up of the investigation, e.g. case control or observational design. Different study designs are not equally good with regard to inferring causality. These differences in study designs translate into different strengths of evidence, and capacity to identify the level of evidence. The study designs are ranked hierarchically according to a level-of-evidence scale (LoE; Fig. 4). The LoE are grouped as follows:

Systematic reviews (LoE1a) are at the top of the evidence hierarchy and provide the most reliable information. They summarize all information gained in several individual studies, have an a priori protocol on design and procedure, and are conducted according to strict guidelines (e.g. Collaboration for Environmental Evidence 2013). Ideally they include quantitative measures, at best a meta-analysis (in the strict sense; see Vetter et al. 2013). Other more conventional reviews (LoE1b) may also include quantitative analysis or are purely qualitative. They both summarize the findings of several studies, but systematic reviews assess the completeness and reproducibility more carefully and try to avoid publication bias by including grey literature (Higgins and Green 2011). The necessary condition for any review is that appropriate individual studies are available. The most reliable individual study is a *study with a reference/control (LoE2).* Typically, these are case-control or before-after control-impact studies (LoE2a). Another useful approach can be the comparison of different approaches or interventions, for example for the valuation of ES where no control exists. Comparing results of different valuation approaches can increase the evidence, if results of both approaches are consistent (LoE2b). *Observational studies (LoE3)* are individual studies without control. These include studies employing inferential and correlative statistics, e.g. testing for the influence of environmental variables on the quantity of an ES (LoE3a). Descriptive studies imply data collection and representation without statistical testing (e.g. data summaries, ordinations, histograms, models with data input). In ES science and conservation these are often surveys or expert elicitations (LoE3b). The lowest level-of-evidence are statements *without underlying data (LoE4)*. These are usually expert opinions, often not distinguishable from randomness (Tetlock 2005). Other statements without underlying data are reasoning based on mechanism and 'first principles': A works according to a certain mechanism, so we expect B to work in the same way. These first principles are not reliable in ecology (Lawton 1999).

After identifying the underlying study design (Fig.4), the implementation of the design needs to be thoroughly assessed through a critical appraisal, to evaluate quality aspects, such as sample size and randomization. A comprehensive list of quality aspects is provided by a quality checklist (Mupepele et al. 2014), facilitating the critical appraisal. The checklist covers questions concerning the methods applied, e.g. "Was the sample size appropriate?" or "Was accuracy/uncertainty measured?". The combination of study design and quality criteria will allow the identification of the level-of-evidence supporting the study result (Fig. 3).

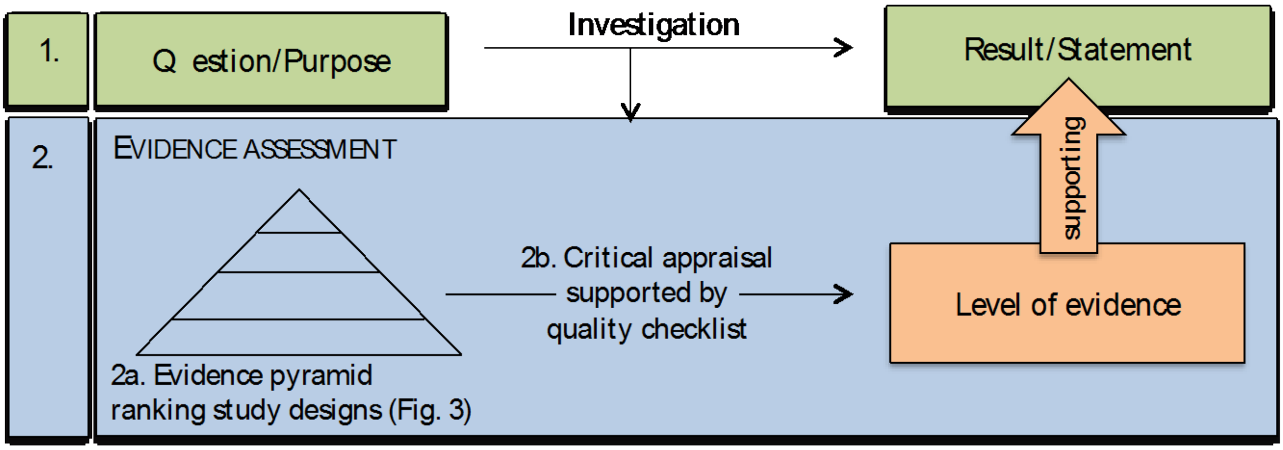

Fig. 3. Schematic procedure of evidence-based practice: 1. Identification of the study question and the outcome, given as result of the study. 2. The assessment of the evidence supporting the result, with help of an evidence hierarchy and a quality checklist.

The evidence-based concept is relevant for various user groups and should be considered as follows:

1. *Scientists conducting their own studies* have to be aware how to achieve the highest possible evidence, particularly during the planning phase. Choosing a study design that provides good evidence and respects quality criteria will substantially increase the potential contribution to our knowledge.
2. *Scientists advising decision-makers* should be aware of the evidence of information they include in their recommendations. Weighting all scientific information equally, or subjectively, runs the risk of overconfidence and bias.
3. *Decision-makers* receiving information from scientists should demand a level-of-evidence statement for the information provided, or should judge themselves the reliability having in mind the evidence-based concept.

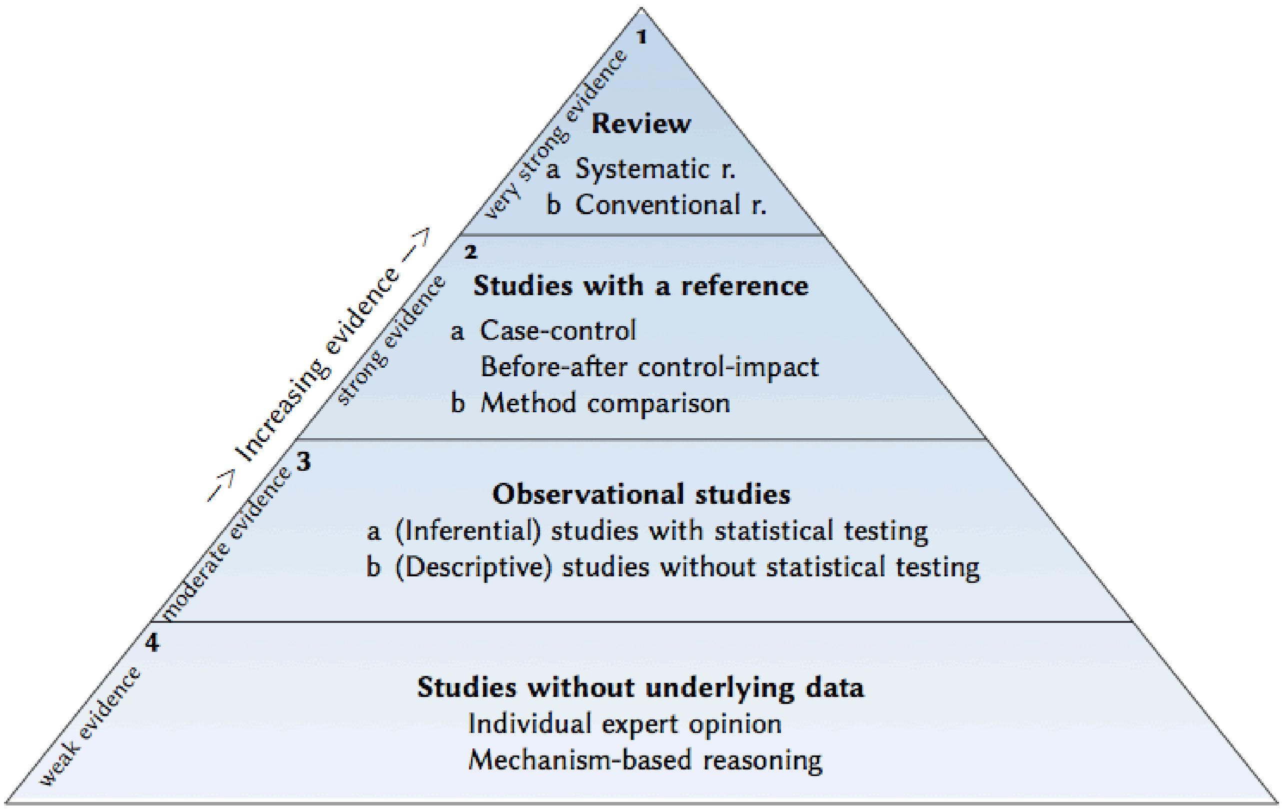

Fig. 4. Level-of-evidence (LoE) hierarchy ranking study designs according to their evidence. LoE1 - LoE4 with internally ranked subcategories a and b.

4 Integration of effectiveness, efficiency and evidence indicator into the OPERAs blueprint

The blueprint protocol is a standardized reporting tool for a consistent comparison, evaluation and synthesis of the ES studies and projects, such as the OPERAs exemplars (see Box 2). The intention to create a standardized framework is based on the realisation that science requires reproducible methodologies and results that can be easily synthesised in order to progress beyond case or site-specific conclusions (Ash et al. 2010, Pullin et al. 2003, 2009). One of the strengths of the blueprint protocol is that it can easily evolve to include effectiveness efficiency and evidence criteria.

Most aspects that are highly relevant for an evidence-based assessment are already included in the current version of the blueprint. In contrast only 53% of the identified effectiveness and efficiency indicators are integrated (Fig. 2). Most frequently are indicators on the experimental set-up indicators for stakeholder engagement and background information on various local conditions. For an adequate estimation of both effectiveness and efficiency some key criteria are still missing. In line with the results of our database review we found similarly neglected indicators in the blueprint protocol which will be discussed and added in future versions. The following indicators are: i) an exhaustive classification system for a consistent foundation of effectiveness and efficiency comparison, ii) detailed information about the impact of the studies, iii) monetary costs and time requirements, iv) capacity building, number of people and champions involved.

Within the blueprint protocol we created our own classification system for categorizing study objectives. It is more detailed than the previous mentioned Daily et al. (2009) framework, however, it neglects some components of the ES concept. For instance only methods for ecological model development and scenario configuration for social and management dimensions are considered. Objectives not covered previously can be added individually with an open-end category, in particular for prescribed overarching objective categories. Therefore, an extension of the objective classification system can be made by adapting it to the Daily et al. (2009) framework. This would be also beneficial for the evidence-based approach, which has shown that it is important to differentiate whether an investigation took part focusing on quantification, valuation, management, governance of ES and/or method development.

To estimate effectiveness the achieved objectives and impact of studies need to be taken into account. Neither indicators for policy uptake nor study-induced alteration in science or environmental changes are considered in the blueprint protocol. On the one hand the blueprint protocol was compiled to gather information of the OPERAs exemplars which are in an ongoing process, thus, only results can be interpreted with respect of the intended impact on the stakeholders or environment. On the other hand to provide a systematic reporting standard beyond OPERAs, indicators that reflect interventions and responses to the key findings of the study should be included.

Monetary cost and time requirements are important indicators in efficiency analysis (Bagstad et al. 2013, Perman et al 2003, Goulder et al. 1999). They constrain the entire analysis framework of an ES study and affect the quality of the findings. Accordingly, transaction costs of the study process and implementation action as well as temporal requirements to achieve the objective of the study or project needs to be considered in the blueprint protocol.

Additionally, capacity building efforts and manpower of a research team strongly affect cost and time issues in a study or project.

The investigation of ES from quantification for valuation and policy work in order to devise payment schemes and management actions requires interdisciplinary qualified groups of expert teams. Manpower crucially affecting cost and time issues in a study or project, in spite of this fact it is not captured. Also the key role of champions with regard to the implementation of science into practice is completely neglected. Therefore, the consideration of the number of people involved in the ES study or project as well as the role of champions for achieving an objective would improve the blueprint regarding the estimation of effectiveness and efficiency.

Box 2. The OPERAs project Blueprint Protocol

What is the Blueprint Protocol?

The blueprint protocol (BP) is a tool to help scientists to carry out an ES assessment in a consistent manner to aid reporting, synthesis and evaluation of the collected data. The development of the protocol was in response to research conducted by Seppelt et al. (2012) and Crossman et al. (2013), calling for a need for consistent and reproducible methodologies within the ecosystem management field. Within OPERAs (Task 2.3.1 and MS 2.4, 2.5, 2.9) the blueprint is designed to ensure consistency of reporting from the research efforts of the Instruments and Exemplar Work Packages and is crucial for the efficient dissemination of information in the Resource Hub.

How does it work?

The current blueprint is accessed via weblink[2] to a google form. It is an online survey with a series of questions, divided into checkboxes, multiple-choice tables and open-end questions that the respondents interactively complete.

The questions cover a range of elements including exemplar study purpose and design, stakeholders, OPERAs team members, OPERAs tool uptake, ES assessed, geographical elements, policy and regulatory aspects, foresight, analysis and monitoring.

When finished, the online form is saved and returned to the BP team (but it is also accessible to any OPERAs team member); collation and analysis will then be carried out.

1 https://docs.google.com/forms/d/1yAL2SDDoR91zLFVuT9avsC23QntfVZSVd6OZGhL7otU/viewform

The blueprint protocol is still in process of optimization and will continue in subsequent editions. Further surveys on what the OPERAs exemplars have done to achieve their goals as well as “What haven’t they done that would have been useful?” and “Why couldn’t they do it?” will be conducted. This helps to both complement the evaluation of effectiveness and efficiency and improve the operationalization of the ES concept in practice. Ultimately, the online blueprints will be available for other scientist from around the world.

5 Summary

Major databases of globally distributed ES examinations capture almost 12.000 studies or projects. Unfortunately, none of these were created to investigate effectiveness and efficiency in ES research. Therefore, databases can provide only limited insights into the estimation of effectiveness and efficiency of ES studies. Nevertheless, indicators for an effectiveness and efficiency analysis could be identified and provide valuable input for further work, especially in OPERAs (but also OpenNESS and other projects). We explored 21 major global ES databases and discovered 541 indicators based on their frequency of use in the database. Accordingly, we structured them into our hierarchical framework and summarized them into 30 indicators of effectiveness and efficiency within four groups (Fig. 2). The highest level of aggregation in our framework resulted from the definition of the terms effectiveness and efficiency. Whereas effectiveness characterizes the (i) *accuracy* and (ii) *completeness* of *achieving an objective*, on the contrary efficiency encompasses the (iii) *effort* of an endeavor to (iv) achieve an objective (Fig. 2).

Primarily, efficiency indicators could be identified in relation to the effort within a study or project, followed by indicators that imply background information of local conditions and drivers. The latter group refers to the category *achieved objectives*. It respects local peculiarities that may lead to implementations of study/project findings and further ensure the comparability of the studies/projects. Third most are effectiveness indicators for achieving an objective, but they occur significant less frequently. Entities on the quality of an ES study or project, i.e. the accuracy and completeness, are the least common indicator. More detailed analysis of disaggregated indicators showed that the majority refer to economic valuation of ES and highlight the continually prevailing position of monetary valuation in ES research. Also indicators represent information on the experimental set-up of the study or project occurred often. Therefore, highly diverse methodological aspects are mentioned, shown by input data requirements, indicators and metrics for ES used as well as general descriptions of methods and tools characteristics. The least represented indicators are those describing monetary costs and time requirements for conducting a study or project, those on needs of manpower and connecting points that advocate scientific results beyond science communities (champions) as well as monitoring issues and indicators on study-induced impact.

Especially missing information on the quality of an ES study or project and a lack of indicators on study-induced impacts require an evidence-based approach that captures the level of evidence by ranking scientific rigor and guides the reporting of evidence-based practice. In order to close this gap we have developed such an approach within OPERAs (Task 2.1.2). Furthermore, it was striking that due to a high thematic heterogeneity between and within the indicator groups, these are hardly comparable. To enhance the comparability a common classification system of study/project objectives is required. Here we proposed the Daily et al. (2009) framework, because it reflects the holistic notion of the ES concept and defines broad categories that help to distinguish between objectives and allows it easily to integrate more detailed classification systems.

Moreover, we discussed how effectiveness and efficiency as well as evidence can be consistently analyzed in future. Therefore, we briefly introduced the OPERAs blueprint protocol (Task 2.3.1 and

MS 2.4, 2.5, 2.9), a standardized reporting tool for a consistent comparison, evaluation and synthesis of the ES studies and projects, and discussed how it could be improved regarding the consideration of effectiveness, efficiency and evidence.

6 References

Armsworth P.R., K.M.A. Chan, G.C. Daily, P.R. Ehrlich, C. Kremen, T.H. Ricketts, M.A. Sanjayan, 2007. Ecosystem service science and the way forward for conservation Conservation Biology, 21 (6), pp. 1383–1384

Ash, N., Blanco, H., Garcia, K., Tomich, T., Vira, B., Brown, C., Zurek, M., 2010. Ecosystems and Human Well-Being: A Manual for Assessment Practitioners. Island Press, Washington, DC.

Abson DJ, von Wehrden H, Baumgärtner S, Fischer J, Hanspach J, Härdtle W, et al. 2014. Ecosystem services as a boundary object for sustainability. Ecological Economics 103(0): 29-37.

Bagstad, K.J., D. Semmens, S. Waage, and R. Winthrop. 2013. A comparative assessment of tools for ecosystem services quantification and valuation. Ecosystem Services 5: 27-39.

Brondizio, E.S., Gatzweiler, F.W., Zografos, C., Kumar, M., 2010. The socio-cultural context of ecosystem and biodiversity valuation. In: Kumar, P. (Ed.), The Economics of Ecosystems and Biodiversity: Ecological and Economic Foundations. Earthscan, London, pp. 150–181.

Carpenter, S.R., Mooney, H.A., Agard, K., Capistrano, D., DeFries, R.S., Díaz, S., Dietz, T., Duraiappah, A.K., Oteng-Yeboah, A., Pereira, H.M., Perrings, C., Reid, W.V., Sarukhan, J., Scholes, R.J., Whyte, A., 2009. Science for managing ecosystem services: beyond the Millennium Ecosystem Assessment. Proc. Natl. Acad. Sci. U.S.A. 106, 1305–1512.

Carvill, Deasley, Guest, Lehane, Martin, O'Mahony, O'Reilly, O'Sulivan, McGloin, McKeown 2012 Review of Effectiveness of SEA in Ireland Key Findings & Recommendations. Environmental Protection Agency, Ireland. ISBN: 978-1-84095-458-6 https://www.agriculture.gov.ie/media/migration/ruralenvironment/environment/seainformation/SeaEffectivenessReviewMain%20Report.pdf

Collaboration for Environmental Evidence. 2013. Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Technical Report MarchEnvironmental Evidence.

COP 2010: COP 10 Documents. UNEP/CBD/COP/10/27. http://www.cbd.int/cop10/doc/

Cranford, M., Parker, C., and Trivedi, M., 2011. Understanding forest bonds, Global Canopy Programme, Oxford, UK

Crossman, N.D., Burkhard, B., Nedkov, S., Willemen, L., Petz, K., Palomo, I., Drakou, E.G., Martín-Lopez, B., McPhearson, T., Boyanova, K., Alkemade, R., Egoh, B., Dunbar, M.B., Maes, J., 2013. A blueprint for mapping and modelling ecosystem services. Ecosyst. Serv. 1–11.

Daily, G.C., Polasky, S., Goldstein, J., Kareiva, P.M., Mooney, H.A., Pejchar, L., Ricketts, T.H., Salzman, J., Shallenberger, R., 2009. Ecosystem services in decision making: time to deliver. Front. Ecol. Environ. 7, 21–28.

Daily G.C., Matson P.A. 2008. Ecosystem services: from theory to implementation Proceedings of the National Academy of Sciences of the United States of America, 105 , pp. 9455–9456.

de Groot, Brander, van der Ploeg, Costanza, Bernard, Braat, Bernardd, Christie, Crossman, Ghermandi, Hein, Hussain, Kumar, McVittie, Portela, Rodriguez, ten Brink, Beukering 2012. Global estimates of the value of ecosystems and their services in monetary unit. Ecosys. Serv. 1, pp. 50–61.

EU 2014: Disclosure of non-financial information: Europe's largest companies to be more transparent on social and environmental issues. http://europa.eu/rapid/press-release_STATEMENT-14-291_en.htm

EU 2013: Growing a Low Carbon, Resource Efficient Economy with a Sustainable Supply of Raw Materials. Horizon 2020-SC5-2015-two-stage. More effective ecosystem restoration in the EU. European Comission, Brussels. http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2195-sc5-07-

2015.html

Gibson CC, Ostrom E, Ahn TK. 2000. The concept of scale and the human dimensions of global change: a survey. Ecological Economics 32(2): 217-239.

Goulder, L.H., Parry I.W.H., Williams III. R.C. 1999. The cost-effectiveness of alternative instruments for environmental protection in a second-best setting J. Pub. Econ., 72, 329–360.

Gregory, R. 2000. Using stakeholder values to make smarter environmental decisions. Environment 42, 34–44.

Gregory, R., McDaniels T., Fields D. 2001. Decision aiding, not dispute resolution: creating insights through structured environmental decisions. J Policy Anal Manag 20, 415–432.

Hanson, C., J. Ranganathan, C. Iceland, and J. Finisdore. 2012. The Corporate Ecosystem Services Review: Guidelines for Identifying Business Risks and Opportunities Arising from Ecosystem Change. Version 2.0. Washington, DC: World Resources Institute.

Hanson, C., Ranganathan, J., Iceland, C., Finisdore, J., 2008. The corporate ecosystem services review: guidelines for identifying business risks and opportunities arising from ecosystem change, World Resources Institute.

Heal, G., 2000. Valuing ecosystem services. Ecosystems 3, 24e30.

Higgins JPT and Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions.

Ingraham M.W., Foster S.G. 2008. The value of ecosystem services provided by the US National Wildlife Refuge System in the contiguous US Ecological economics 67(4):608 -618.

IPBES 2013: Critical review of the assessment landscape for biodiversity and ecosystem services. IPBES/1/INF/8. http://www.ipbes.net/images/IPBES_1_INF_8_En.pdf

ISO 5725, -1:1994(en) Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions. https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en

Jakeman AJ, Letcher RA, Norton JP. 2006. Ten iterative steps in development and evaluation of environmental models. Environ Modell Softw 21(5): 602-614.

Jacobs, M., 1997. Environmental valuation, deliberative democracy and public decision-making institutions. In: Foster, J. (Ed.), Valuing Nature? Economics, Ethics and Environment. Routledge, London, pp. 211–231.

Kandziora M, Burkhard B, Müller F., 2013. Interactions of ecosystem properties, ecosystem integrity and ecosystem service indicators- theoretical matrix exercise. Ecol Ind 28:54–78

Kareiva P., M. Marvier, 2003. Conserving biodiversity coldspots - Recent calls to direct conservation funding to the world's biodiversity hotspots may be bad investment advise. Am. Sci. 91, 344.

Keynes, J. M. 1937 The general theory. Quart. J. Econ. 51, 209–233. (doi:10.2307/1882087)

Kettunen, M., Bassi, S., Gantioler, S., ten Brink, P., 2009. Assessing socio-economic benefits of natura 2000 - a toolkit for practitioners.

Kosmus, M., Renner, I. and Ullrich, S., 2012. Integrating Ecosystem Services into Development Planning: A stepwise approach for practitioners based on the TEEB approach, GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit) GmbH, Eschborn, Germany

Laurans Y, Rankovic A, Bille R, Pirard R, Mermet L. 2013. Use of ecosystem services economic valuation for decision making: Questioning a literature blindspot. J Environ Manage 119: 208-219.

Lawton JH. 1999. Are There General Laws in Ecology? Oikos 84: 177–192.

Lorenzoni, I., A. Jordan, M. Hulme, R.K. Turner and T. O'Riordan, 2000: A co-evolutionary approach to climate change impact assessment. Part I. Integrating socio-economic and climate change scenarios. Global Environ. Chang., 10, 57-68.

MARRS, RH., GALTRESS, K., TONG, C., BLACKBIRD, SJ., HEYES, TJ., PAKEMAN, RJ., LE DUC, MG., 2007. Conflicts between competing conservation goals, biodiversity or ecosystem services: quantification of element losses and species recruitment using a moorland-bracken model system. Journal of Environmental Management, vol. 85, p. 1034-1047. PMid:17207569. http://dx.doi.org/10.1016/j.jenvman.2006.11.011

Martínez-Alier, J. 2002. The environmentalism of the poor. Edward Elgar, London.

Martin-Lopez B, Gomez-Baggethun E, Garcia-Llorente M, Montes C. 2013. Trade-offs across value-domains in ecosystem services assessment. Ecological Indicators 37: 220-228.

Mupepele A-C and Dormann CF. 2014. Assessing the evidence of ecosystem services studies: a framework and its application. BioRxiv: http://biorxiv.org/lookup/doi/10.1101/010140

Nakhooda S., Forstater M.. 2013: The effectiveness of climate finance: a review of the Global Environment Facility. Overseas Development Institute. London. ISSN (online): 1759-2917. http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8632.pdf

Nilsson J., Bergstrom S. 1995. Indicators for the assessment of ecological and economic consequences of municipal policies for resource use. Ecological Economics, 14: 175–184

Oxfam 2012. Effectiveness reviews. Evaluation Reports. http://policy-practice.oxfam.org.uk/our-work/methods-approaches/project-effectiveness-reviews

Olschewski R., Klein A.. 2011. Ecosystem services between sustainability and efficiency. Sustainability: Science, Practice, & Policy 7(1):69-73. Published online Mar 24, 2011. http://www.google.com/archives/vol7iss1/communityessay.olschewski.html

Paavola, J., and K. Hubacek. 2013. Ecosystem services, governance, and stakeholder participation: an introduction. Ecology and Society 18(4): 42. http://dx.doi.org/10.5751/ES-06019-180442

Perman, R., Y. Ma, J. McGilvray, M. 2003. Common. Natural Resources and Environmental Economics, Pearson Higher Education Addison Wesley

Phillips C., Thompson, G. 2009: What is cost-effectiveness? Hayward Medical Communications, Hayward Group Ltd. http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Cost-effect.pdf

Phoenix G.K., Johnson D., Grime J.P., Booth R.E. 2008. Sustaining ecosystem services in ancient limestone grassland: importance of major component plants and community composition. Journal of ecology 96(5): 894 -902

Piñeros- Garcet, J.D., van Bol, V., Pussemier, L., 2010. HArmonised environmental Indicators for pesticides Risks, HAIR, contract number SSPE-CT-2003-501997. http://www.rivm.nl/rvs/Images/HAIR_AGGREGATION_OF_INDICATORS_tcm35-40199.pdf

Pullin AS and Knight TM. 2003. Support for decision making in conservation practice: an evidence-based approach. Journal for Nature Conservation 11: 83–90.

Pullin AS and Knight TM. 2009. Doing more good than harm - Building an evidence-base for conservation and environmental management. Biological Conservation 142: 931–4.

Ranganathan, J., Raudsepp-Hearne, C., Lucas, N., Irwin, F., Zurek, M., Bennet, K., Ash, N., West, P., 2008. Ecosystem Services - A Guide for Decision Makers.

Reed, M. S., K. Hubacek, A. Bonn, T. P. Burt, J. Holden, L. C. Stringer, N. Beharry-Borg, S. Buckmaster, D. Chapman, P. J. Chapman, G. D. Clay, S. J. Cornell, A. J. Dougill, A. C. Evely, E. D. G. Fraser, N. Jin, B. J. Irvine, M. J. Kirkby, W. E. Kunin, C. Prell, C. H. Quinn, B. Slee, S. Stagl, M. Termansen, S. Thorp, and F. Worrall. 2013: Anticipating and managing future trade-offs and complementarities between ecosystem services. Ecology and Society 18(1): 5. http://dx.doi.org/10.5751/ES-04924-180105

Ruckelshaus M, McKenzie E, Tallis H, Guerry A, Daily G, Kareiva P, et al. 2013. Notes from the field: Lessons learned from using ecosystem service approaches to inform real-world decisions. Ecological

Economics(0).

Rumsfeld, D. 2002. Defense.gov news transcript. http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636

Schroeter, M., van der Zanden, E. H., van Oudenhoven, A. P. E., Remme, R. P., Serna-Chavez, H. M., de Groot, R. S., and Opdam, P., 2014. Ecosystem services as a contested concept: a synthesis of critique and counter-arguments. Conservation Letter, doi: 10.1111/conl.12091

Seppelt, R., Fath, B., Burkhard, B., Fisher, J.L., Grêt-regamey, A., Lautenbach, S., Pert, P., Hotes, S., Spangenberg, J., Verburg, P.H., Oudenhoven, A.P.E.E. Van, 2012. Form follows function? Proposing a blueprint for ecosystem service assessments based on reviews and case studies. Ecol. Indic. 21, 145–154.

Seppelt R, Dormann CF, Eppink FV, Lautenbach S, Schmidt S. 2011. A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. J Appl Ecol 48(3): 630-636.

Spiegelhalter DJ, Riesch H. 2011. Don't know, can't know: embracing deeper uncertainties when analysing risks. Philos Trans R Soc A-Math Phys Eng Sci 369(1956): 4730-4750.

Sprinz, D. F. 2000. Research on the Effectiveness of International Environmental Regimes: A Review of the State of the Art. Paper Prepared for the Final Conference of the EU Concerted Action on Regime Effectiveness, Institut D'educació Contínua (IDEC), 09 – 12 November 2000, Barcelona. http://www.uni-potsdam.de/u/sprinz/doc/sprinz.effectiveness.pdf

Stolton, S. and Dudley, N., 2008. The Protected Areas Benefits Assessment Tool, WWF International, Gland, Switzerland, ISBN: 978-2-88085-284-9.

Tetlock PE. 2005. Expert Political Judgment: How Good Is It? How Can We Know? Princeton University Press.

UK NEA 2011. The UK National Ecosystem Assessment: Synthesis of the Key Findings. UNEP-WCMC, Cambridge.

van Beukering PJH, Cesar HSJ, Janssen MA. 2003. Economic valuation of the Leuser National Park on Sumatra, Indonesia. Ecological Economics 44(1): 43-62.

van Wilgen B. W., Le Maitre D. C., Cowling R. M. 1998. Ecosystem services, efficiency, sustainability and equity: South Africa's Working for Water Programme. Trends in ecology & evolution 13(9):378 -378.

van Wilgen B. W., Reyers B., Le Maitre D.C., Richardson D.M., Schonegevel L. 2008. A blome-scale assessment of the impact of invasive alien plants on ecosystem services in South Africa. Journal of environmental management [0301-4797],vol:89, iss:4, pg:336 -349Villa, F., K.J. Bagstad, B. Voigt, G.W. Johnson, R. Portela, M. Honzak, and D. Batker. in press. A methodology for adaptable and robust ecosystem services assessment. Forthcoming in: PLOS ONE.

Vatn, A., 2005. Rationality, institutions and environmental policy. Ecol. Econ. 55, 203–217.

Vatn, A., Bromley, D., 1994. Choices without prices without apologies. J. Environ. Econ. Manage. 26, 129–148.

Vetter D., Rücker G., and Storch I. 2013: Meta-analysis: A need for well-defined usage in ecology and conservation biology. Ecosphere 4: art74.

Villa F, Bagstad KJ, Voigt B, Johnson GW, Portela R, Honzak M, et al. 2014. A Methodology for Adaptable and Robust Ecosystem Services Assessment. PLoS One 9(3).

Vogel A. 2012: Aligning Water Fund Investment Prioritization in Latin America: General Guidance http://www.naturalcapitalproject.org/pubs/Water_Fund_Prioritization_Guidance_Document.pdf

Walker W.E., P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, M.P. Krayer von Krauss 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support Integrated Assessment, 4, pp. 5–18

Wegner, G., Pascual, U., 2011. Cost–benefit analysis in the context of ecosystem services for human well-being: a multidisciplinary critique. Global Environ. Change 21, 492–504

Whitlock, Lopez, Chang , Helfand , Eder, Floyd. 2009. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. J Clin Epidemiol. 2010 May;63(5):491-501. doi: 10.1016/j.jclinepi.2009.03.008. Epub 2009 Jun 21. http://effectivehealthcare.ahrq.gov/ehc/products/60/318/20090427IdenttifyingTopics.pdf

Wynne, B. 1992. Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm. Global Environ. Change 2, 111–127. (doi:10.1016/0959-3780(92)90017-2)

Zhao B, Kreuter U, Li B, Ma ZJ, Chen JK, Nakagoshi N. 2004. An ecosystem service value assessment of land-use change on Chongming Island, China Land Use Policy 21(2):139 -148

Zscheischler J, Rogga S, Weith T. 2014. Experiences with Transdisciplinary Research. Syst Res Behav Sci 31(6): 751-756.

Appendix

Appendix 1. References of database used to identify effectiveness and efficiency indicator.

Database name	Source	Web host
Beneficial Use Values Database (BUVD)	Official BUVD website: buvd.ucdavis.edu	University of California, Davis, Department of Agricultural and Resource Economics; Project coordinators: Douglas M. Larson
Environmental Valuation & Cost-Benefit News (EVCBN)	Official EVCBN website: http://www.envirovaluation.org/	Cost Benefit Group (CBG), formerly Damage Valuation Associates (DVA) http://www.costbenefitgroup.com/mission.htm
Sub-Global Assessments database (SGA)	Website of the Millennium Ecosystem Assessment: http://www.unep.org/maweb/en/Multiscale.aspx	Sub-Global Working Group; matthew.dixon@unep-wcmc.org
Catalogue of Assessments on Biodiversity and Ecosystem Services (CA)	Website of Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES): http://catalog.ipbes.net/	United Nations Environment Programme (UNEP)
IIED Watershed Markets	Official IIED Website: http://www.watershedmarkets.org/index.html	International Institute for Environment and Development (IIED):
TEEB – The Economics of Ecosystems and Biodiversity cases	Official TEEB website: http://www.teebweb.org/resources/case-studies/	UNEP TEEB Office and Sharon Oseku-Frainier (TEEB Communication)
Payment for Ecosystem Services Database (PES)	Website of the Organization of American States (OAS): http://www.oas.org/dsd/PES/Database.htm#	Department of Sustainable Development (DSD)
Review of Externality Data (RED)	Official RED website: http://www.isis-it.net/red/	European Comission, Energy, Environment and Sustainable Development Programme of the Directorate General for Research
Ecosystem Services Bibliography (ESB)	Official ESB blog: http://blog.lib.umn.edu/polasky/ecosystem/	University Of Minnesota Libraries
ARtificial Intelligence for Ecosystem Services (ARIES) case studies	Official ARIES website: http://www.ariesonline.org/	University of Vermont, Gund Institute for Ecological Economics: Ecoinformatics Collaboratory
Environmental Valuation Database (ENVALUE)	Official ENVALUE website: http://www.environment.nsw.gov.au/envalueapp/Default.asp?ordertype=MEDIUM	New South Wales Environmental Protection Authority, Department of Environment, Climate Change and Water
Marine Ecosystem Service Partnership (MESP)	Official MESP website: http://www.marineecosystemservices.org/explore	Duke University, Nicholas Institute for Environmental Policy Solutions

Environmental & Recreational (Non-Market) Values - Valuation Studies Search from National Ocean Economics Program (NOEP)	Official NOEP website: http://www.oceaneconomics.org/nonmarket/NMsearch2.asp	U.S. National Oceanic and Atmospheric Administration
ZEN database	Publication: Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S., Schmidt, S., 2011. A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. J. Appl. Ecol. 48, 630–636.	Helmholtz Centre for Environmental Research – UFZ, Department Computational Landscape Ecology
Ecosystem Service Valuation Database (ESVD)	Webpage of the Ecosystem Service Partnership (ESP): http://www.es-partnership.org/esp/80763/5/0/50	Foundation for Sustainable Development
Vihervaara et al. 2010	Publication: Vihervaara P, Ronka M, Walls M: Trends in ecosystem service research: early steps and current drivers. Ambio 2010 39:314-324.	University of Turku, Department of Biology
Liu et al. 2010	Publication: Liu S, Costanza R, Farber S, Troy A: Valuing ecosystems services: theory, practice, and the need for aransdisciplinary synthesis. Ann N Y Acad Sci 2010, 1185:54-78.	University of Vermont, Gund Institute of Ecological Economics and Rubenstein School of Environment and Natural Resources
de Bello et al. 2010	Publication: de Bello F, Lavorel S, Díaz S, Harrington R, Cornelissen JHC, Bardgett RD, Berg MP, Cipriotti P, Feld CK, Hering D et al.: Towards an assessment of multiple ecosystem processes and services via functional traits. Biodivers Conserv 2010, 19:2873-2893.	National Centre for Scientific Research, Grenoble
Goldman et al. 2008	Official website of the Natural Capital Project: http://www.naturalcapitalproject.org/database.html	Stanford University, Department of Biological Sciences
Ecosystem Service Indicator Database (ESID)	Official website of the ESID: http://www.esindicators.org/indicators_overview	World Resources Institute
Environmental Valuation Reference Inventory (EVRI)	Official website for EVRI: https://www.evri.ca/Global/Splash.aspx	Environment Canada on behalf of other EVRI member institutions

Appendix 2. Explanation of effectiveness and efficiency indicator used in major global ES databases. The column "Examples from databases" show only a subset of indicators that were actually considered in this analysis.

Indicator name	Explanation	Examples from databases
costs	The use of data gathering, involvement of experts, application of tools and methods to reach the objectives of the study or project based on market prices.	funding involved
time	Temporal requirements to achieve the objectives of the study or project.	project length; year assessment started/finished; periodicity of assessment; if repeated, how frequently, period assessed
monetarization of ES	Indicator that describe the effort, i.e. use of economic valuation method to appraise ES in monetary units.	type of monetary valuation methods; values; valuation years; links valuation physical impact
experimental set-up	Aggregation of different indicators that describe the effort (complexity) and tools to measure ES.	tools and approach used; input required; indicators/units used
on-the-ground implementation	Specific activities, mechanism and tools that were used to enable implementation in practice.	specific recommendations; specific activities to achieve the goals (financial instruments, easement used, institutional/legal tools)
policy analysis	Indicator that describe political/legal needs in order to be able to institute project actions.	Was a policy analysis done for the project? Use of policy analysis? Politically/legally needs in order to institute project actions
uncertainties	Represents the accuracy of the findings.	general uncertainty; reviewed; validated; quality of results
ES-subtype	Indirect effectiveness and efficiency indicator on which ES subtypes are investigated. Can be used to structure ES studies/projects and estimate the effort of the analysis.	food: beef, fish; extreme events: flood prevention, storm protection etc.
exact investigation area	Indirect effectiveness and efficiency indicator on which area is investigated with which spatial resolution.	Location name; project ecoregion; Specific geographic locations of application; Receiving Environment
project description	General description on indicators for methods used as well as drivers, pressures and general background information on local conditions.	Abstract and project description
data source	Indirect effectiveness and efficiency indicator on which input data is used to estimate the effort and completeness of the ES examination (under consideration of processing and output).	primary; secondary data; references
size of investigation area	Indirect effectiveness and efficiency indicator on how big the area is that was investigated by the study/project. The indicator can be used to ensure the comparability of studies.	service area in sqkm
scenario	Indirect effectiveness and efficiency indicator on which kind of scenarios are conducted. Can be understood as an indicator for the on-the-ground implementation (see above; Reed	Scenario analysis; Tools and approaches used in the assessment

	et al. 2013) or as a further indication to estimate efforts of the examination.	
interdependences of ES	Indirect effectiveness and efficiency indicator that represents whether trade-offs or synergies between different ES are considered. It indicates the accuracy and efficiency.	ES in isolation examined; combination of traits
expertise & capacity building	The background and know-how of people involved in a study affect the proceeding of the study, thus, the efficiency (Reed et al. 2013).	Research institute/group; organization; Capacity building needs identified during the assessment; How have gaps in capacity been communicated to the different stakeholders
stakeholder engagement	Stakeholder engagement help to ensure the credibility, relevance and legitimacy of a study/project, and increases the extent to which findings are reflected in decision making (Gregory et al. 2001, Gregory 2000).	Are stakeholder engaged? Does the study try to engage the community? If yes, using what mechanisms? How do they communicate project goals?
champion	Experts that link science and policy to implement scientific results in decision making processes.	Project implementation (In order to achieve project goals/targets what are the major conservation actions/on the ground activities occurring in the project area (Who is implementing these activities)?); broker involved; facilitator
manpower	Number of experts that were involved in the study/project	The number of people directly involved in the assessment process
Environmental change	Implementation of study/project findings in nature	Number of seedlings planted, Number of acres restored; Change in flood risk
monitoring	Systematic measurement or observing of processes and indicator resulting from study/project findings.	Compliance and/or performance monitoring; indicators; What is being monitored? How is it being monitored? In how many locations is the monitoring occurring? With what frequency? When did this start? Who (what organization) is collecting data? Who analyzes data?
uptake	Implementation of study/project findings in decision making or society in general	interventions and response to the key findings of the study/project; policy impact; legislation issues; capacity building by the assessment
lessons learned	Consequences and take home message of the study/project without necessarily being implemented in practice. May include indications on how to make ES examination more effective and efficient in future.	Consequences; challenges and lessons learned; What were the key challenges in creating the project? The project process?
objective	Effectiveness and efficiency can only be analyzed relatively to the objective, i.e. desired aim and achieved goal (result) of the study/project.	purpose and objectives; ecosystem service; habitat; species; socio-economic targets; mandate for the assessment
results	Effectiveness and efficiency can only be analyzed relatively to the objective, i.e. desired aim and achieved goal (result) of the study/project.	key results of the studies/projects

winners/losers	Indirect effectiveness and efficiency indicator that identifies winner and loser. Link study/project results to people that are affected is crucial to increase relevance for policy making (Paavola et al. 2013).	Buyer/Investor; seller; stakeholder (supply, demand, facilitator, intermediary)
outreach	Indirect effectiveness and efficiency indicator on multiple ways of communicating research results not only through publishing in research outlets but also through broadcasting documentaries etc. (Reed et al. 2013)	assessment outputs: website, report, communication material, journal publication, training materials
drivers	Indirect effectiveness and efficiency indicator on which driver cause the purpose of the study. The level of intensity may affect the level of uptake.	What are the major threats/main threats to the project area? Extent of Environmental Change; Drivers of change / Driver of Ecosystem Change; What was the problem?
various local conditions	Indirect effectiveness and efficiency indicator on which driver and local structures and processes cause the purpose of the study. The local conditions may affect the level of uptake.	background information of local conditions and description why indicators is important for the region; protected area; income group; pop density; World Bank group;
country	Indirect effectiveness and efficiency indicator that represents the political division in which the study/project took place.	Country or countries covered; Site Description
biome	Indirect effectiveness and efficiency indicator that represents the biophysical area in which the study/project took place.	Biome; System; Ecosystem