

ENVIRONMENT DIRECTORATE
ENVIRONMENT POLICY COMMITTEE

Working Party on Integrating Environmental and Economic Policies

Measuring Environmental Compliance: Designing Analytically Sound and Policy-Relevant Indicators

This document reviews innovative approaches used by OECD countries in the design of intermediate outcome indicators of compliance. It aims at helping environmental enforcement authorities (EEAs) to make their outcome performance measurement more effective, efficient, and open to potential benchmarking across different sub-national EEAs in decentralised systems of environmental governance. The paper was produced under output area 2.3.3.3 of the Programme of Work and Budget for 2013-2014.

The document was presented and discussed at the WPIEEP meeting on 18-19 November 2014, and written comments have been received following the meeting. This is a revised draft incorporating the delegates' comments.

For additional information, please contact Mr Eugene Mazur, Environmental Performance and Information Division, OECD Environment Directorate, tel. +33 1 45 24 76 92, e-mail eugene.mazur@oecd.org.

JT03368650

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	4
1. INTRODUCTION	6
1.1 Background.....	6
1.2 Objective and scope	7
1.3 Methodology	7
2. IMPROVING THE MEASUREMENT OF COMPLIANCE	9
2.1 Analytical constraints of compliance measurement.....	9
2.2 Categorisation of non-compliance	10
2.2.1 Degrees of environmental harm	10
2.2.2 Differentiation of licence conditions.....	11
2.3 Measuring overall compliance performance of regulated entities	12
2.3.1 Qualitative compliance rating	12
2.3.2 Linkages between compliance assessment and risk assessment	13
2.4 Possible compliance targets and indicators.....	15
3. COMPOSITE INDICATORS	18
3.1 Advantages and disadvantages of aggregate measures.....	18
3.2 Experience with compliance-related indices.....	18
3.2.1 Pollution release indicators	18
3.2.2 Facility-specific compliance indicators.....	19
3.2.3 Agency-wide performance indices	20
4. CROSS-JURISDICTIONAL COMPARABILITY OF OUTCOME INDICATORS OF COMPLIANCE ASSURANCE	21
4.1 Drivers and constraints of comparability and benchmarking	21
4.2 Potentially comparable outcome measures	22
4.2.1 Measures of serious non-compliance	22
4.2.2 Occurrence of significant pollution incidents.....	22
4.2.3 Pollution release indicators	23
4.3.2 Monetary value of complying actions	23
5. OPTIMISATION OF A SUITE OF INDICATORS	24
5.1 Reducing the data management burden	24
5.2 Adapting indicators to different audiences	25
5.3 Challenges of quantifying costs and benefits of performance measurement	26
BIBLIOGRAPHY	27
ANNEX. PERFORMANCE EVALUATION OF ENVIRONMENTAL ENFORCEMENT IN CHINA: FIRST LESSONS LEARNED	29
1. Objectives of the pilot programme.....	29

2. Pilot programme participants and key milestones	29
3. Reference performance index	30
4. Reported implementation to-date.....	31
5. Key implementation challenges	31
6. Lessons learned.....	32
7. Next steps in improving the performance measurement system.....	33

Tables

Table 1. Severity of consequence of non-compliance: Classification by EPA Ireland	11
Table 2. Scottish EPA's compliance matrix.....	13

Figures

Figure 1. Compliance with permits in England by performance band, 2006-2012	17
---	----

Boxes

Scoring facilities' enforcement record in Ireland	14
Compliance rating in the Opra scheme in England.....	15
Performance data management at EPA Victoria.....	25

EXECUTIVE SUMMARY

1. This paper aims to help environmental enforcement authorities (EEAs) to make measurement of their performance in securing compliance with environmental requirements more effective, efficient, and open to benchmarking. It builds on the earlier analysis of the experience of OECD countries in the design and implementation of several types of intermediate outcome indicators – measures of changes in the regulated community’s knowledge or behaviour as a result of EEAs’ environmental compliance assurance efforts.

2. The present work looks at a number of possible ways to improve the analytical soundness and policy relevance of outcome indicators of compliance assurance, particularly by:

- Designing compliance indicators to reflect the seriousness of environmental offences and/or the overall compliance performance of regulated entities, including composite measures;
- Developing comparable indicators to benchmark across sub-national jurisdictions; and
- Weighing the costs of tracking a certain performance measure against the benefits it brings to the organisation.

Improving the measurement of compliance

3. One way that EEAs could measure performance is by assessing compliance rates, that is the share of regulated entities complying with environmental requirements. However, determining statistically representative compliance rates is almost impossible because EEAs increasingly target their compliance assurance activities according to the risks involved. In addition, a simple binary – yes/no – approach to measuring compliance does not distinguish between relatively minor and serious violations.

4. In keeping with a risk-based approach, it is advisable to develop compliance indicators that aim to classify non-compliance in terms of its significance for human health or the environment. This may involve categorising the extent to which the offence may result, or has resulted, in harm to the environment, human health or safety (e.g. to assess the occurrence of violations with a risk of serious impact). It is also possible to grade the severity of non-compliance based on the degree to which a certain (type of) permit or licence condition has been exceeded (e.g. minor or gross breaches, with respective operational definitions). In addition, the number of significant pollution incidents (at both permitted and non-permitted facilities) can be used as a surrogate measure of environmental impact.

5. Another good practice approach is to ***assess the compliance behaviour of individual regulated entities in a more comprehensive manner***. One example of this approach is establishing a compliance rating scheme with several performance “bands”. Individual facilities are assessed using quantitative and qualitative indicators of non-compliance and then assigned to a particular band. The EEA can then measure the share of regulated entities in each band and their movement across the bands over time. Measuring trends over time enables an assessment to be made of the impacts of EEA’s activities. ***Compliance rating can also be conveniently integrated into the general assessment of environmental risk of each regulated entity.***

6. Compliance rating and environmental risk scoring schemes (pioneered in the UK) also represent a feasible way to use composite indicators in measuring compliance assurance outcomes. Such composite indicators aggregate and weigh a set of compliance-related parameters to arrive at one measure.

7. Generally, the potential communication benefits of composite indices are outweighed by the costs of collecting and processing large volumes of information, and aggregating it in a methodologically robust manner. Moreover, it is often difficult to interpret policy implications of broad composite indicators and to use them for strategic planning purposes. However, where composite indicators do not aspire to describe environmental outcomes or overall agency performance but rather aggregate different factors of the same intermediate outcome (e.g. via compliance bands), they can be effectively built into the existing strategic planning and operational practice. For example, when performance bands are used, this can be done by setting and monitoring the achievement of targets to reduce the share of regulated entities in the worst performance band.

Cross-jurisdictional comparability

8. In decentralised governance systems, it is important to develop a small set of comparable outcome indicators. This can facilitate benchmarking of performance across sub-national EEAs, promote a level playing field in environmental regulation and compliance assurance, and foster peer learning.

9. In developing comparable indicators, account should be taken of the differences in the definition and size of the regulated community. One approach is to ***limit the scope of such indicators to permitted or licensed facilities*** which are relatively easy to define. From a comparability perspective, it is also advisable to ***use ratios or percentages rather than absolute numbers***, e.g. to measure non-compliance per number of permitted facilities.

10. Rates of serious non-compliance, of significant pollution incidents, and compliance scores and bands, appear to be the most promising comparable measures. If there are differences in the definitions of “severity”, those can be offset by comparing trends in similar indicators over a number of years. It may also be possible to conduct inter-agency comparison with the help of certain indicators of pollution releases or improved environmental management practices, although they have greater practical limitations.

Optimisation of a suite of indicators

11. The performance measurement system should be supported by adequate performance measurement and data collection and reporting mechanisms. Spending staff time and resources on gathering and processing data that are not essential for the agency’s management process may be excessively burdensome for the organisation. The pressure to reduce operating costs becomes a driver for making the performance measurement system leaner and more meaningful.

12. EEAs require indicators for different purposes and audiences. It may be useful to differentiate between a larger “dashboard” of measures for managers within the agency and smaller “scorecard” for policy makers and the public. The EEA should ***review the policy relevance of individual measures to check whether they fit their particular purpose and audience.***

13. In optimising the suite of compliance assurance indicators, the EEA would ideally want to weigh the costs and benefits of tracking individual performance measures. However, in practice EEAs find it difficult to monetise even the benefits of their core activities, let alone those of performance measurement. As a result, EEAs would tend to rely on compliance assurance indicators for which data are readily available. Among outcome measures, ***pollution release and pollution incident indicators based on routine emission reporting and incident notification as well as compliance measures derived from inspection data or operators’ self-reporting are likely to be most cost-effective.***

1. INTRODUCTION

1.1 Background

14. Periodic evaluation of environmental compliance assurance¹ programmes serves a number of essential purposes, including progress assessment, internal and external accountability, and contribution to the deterrence of non-compliance.

15. Very few OECD countries have developed quantitative indicators to measure the outcomes of implementing environmental laws and regulations, and there are no indicators yet that would allow international comparisons in this area. This creates an information gap on issues relevant for both environmental and economic policies, for instance, with respect to the size of unlawful economic gains as a result of non-compliance. The OECD project on outcome performance measures of environmental compliance assurance aims to address this problem, thereby increasing confidence in the consistent and systematic application of environmental requirements.

16. In the first phase of this work (2009-2010), the OECD analysed the experience of ten member countries² in the design and implementation of quantitative indicators used to assess the outcomes of environmental enforcement authorities' efforts to ensure compliance with pollution prevention and control regulations. That study³ considered the following six types of outcome performance measures, which aim to assess improvements either in the behaviour of the regulated community (intermediate outcomes) or in environmental conditions (final outcomes):

- Compliance rates;
- Measures of recidivism and duration of non-compliance;
- Pollution release indicators;
- Indicators of improved environmental management practices and reduced risk;
- Measures of effectiveness of compliance assurance; and
- Environmental quality indicators.

17. Based on the OECD criteria for the evaluation of environmental indicators – measurability, analytical soundness and policy relevance – the study identified key challenges for developing and using

¹ Environmental compliance assurance is the application of all available instruments (principally, those of compliance promotion, compliance monitoring, and enforcement against violations) aimed at influencing the behaviour of regulated entities so that they comply with regulatory requirements.

² Australia, Belgium (Flanders), Canada, Denmark, Ireland, the Netherlands, Poland, Switzerland, the United Kingdom (England and Wales) and the United States.

³ <http://dx.doi.org/10.1787/5kmd9j75cf44-en>

specific categories of compliance assurance outcome indicators and suggested several ways to improve their effectiveness.

18. The review of a “toolbox” of existing outcome indicators and the analysis of their respective strengths and weaknesses suggests that it is not possible to identify a “best practice” approach or a universal optimal set of indicators. The functionality of individual outcome measures ultimately depends on their purpose and suitability for joint analysis with the enforcement authority’s resource (input) and activity (output) indicators. The first phase of the work also identified the importance of associating outcome indicators with time-specific targets as a means of integrating the strategic planning and performance management processes.

1.2 Objective and scope

19. The primary *objective* of this paper is to help environmental enforcement authorities in OECD countries to adequately measure the effectiveness of their efforts and enable policy makers and the public to see the actual impact of their programmes. As the experience of implementing compliance assurance indicators grows, deeper cross-country analysis of several specific issues and other collaborative efforts are valuable in facilitating the exchange of good practices in this area.

20. In light of guidance provided by OECD countries that participated in the earlier work, the second phase of the project (2013-2014) examined the following issues, reflected in the structure of this report:

- Classifying and measuring of non-compliance based on the degree of its environmental impact;
- Using composite indices and weighting to characterise compliance and pollution reduction outcomes;
- Feasibility of developing a limited number of comparable outcome measures to track compliance with similar environmental regulatory requirements in different sub-national jurisdictions (in decentralised systems of environmental governance) or internationally (for example, in the context of European Union Directives); and
- Optimising the size of an environmental enforcement authority’s suite of outcome performance measures from the cost efficiency perspective.

21. The study focuses in particular on *intermediate outcome indicators*, which characterise changes in compliance knowledge and behaviour of the regulated community. They may cover greater understanding by regulated entities of how to comply with environmental requirements, improved environmental management (adoption of best practices), reduced environmental impact (e.g. pollution releases or accidents) and risk, or increased compliance.

1.3 Methodology

22. The OECD Secretariat worked in close collaboration with the Environment Protection Authorities of the states of South Australia and Victoria (Australia), the Environmental Protection Agency of Ireland, the Scottish Environmental Protection Authority (UK) and the Environment Agency (England, UK). The Ministry of Environmental Protection of the People’s Republic of China was an observer of the project.

23. Experts from these environmental enforcement authorities (EEAs) engaged in dialogue, facilitated by the Secretariat, on the project’s key substantive issues. This dialogue, conducted via the

project website and e-mail, combined exchanging each agency's experience and brainstorming on ways to advance the analysis in areas where there is little practical know-how across the community of environmental regulators.

24. Mr Eugene Mazur of the OECD Environment Directorate is the principal author of this paper. Mr Qin Hu of Environmental Defense Fund China contributed the annex on China's experience in designing a composite index to evaluate the performance of environmental enforcement authorities. The comments and suggestions offered by experts from all the participating countries as well as by Mr Brendan Gillespie, Mr Shardul Agrawala and other colleagues of the OECD Environment Directorate are gratefully acknowledged.

2. IMPROVING THE MEASUREMENT OF COMPLIANCE

2.1 Analytical constraints of compliance measurement

25. Compliance rates are used to describe changes in regulated entities' compliance status, which makes them a key measure of direct impact of compliance assurance activities. A compliance rate can be generally defined as a percentage of a regulated universe (or some portion of it) complying with all or certain specific environmental regulatory requirements over the reporting period⁴.

26. The vast majority of compliance rates are calculated based on the ratio between the number of facilities with violations and the number of facilities inspected, notwithstanding some variations in definitions. These inspections are usually conducted at far less than half of the facilities in the regulated community (given limited governmental resources), and most of them are not conducted randomly but deliberately target high-risk facilities and respond to incidents or complaints. This means that the inspected populations are not representative, the resulting compliance rates are not statistically valid, and one cannot generalise the compliance status of the uninspected facilities, i.e. the sector as a whole, from these rates. Compliance rates calculated based on targeted inspections are in reality a kind of "hit rate" measuring whether those inspections are succeeding at identifying violations. A lower compliance rate may mean that the agency is simply doing a better job of detecting violations (Mazur, 2010).

27. Producing representative, statistically valid rates⁵ based on inspections is difficult due to the limited number of inspections that can be conducted with available resources, and the growing need to target those inspections at higher-risk regulated entities. In a way, demand for better outcome performance management comes here into conflict with the major trend of risk-based targeting of compliance monitoring. Risk-based targeting of inspections is meant to find more non-compliance than the presumed average across the regulated community, thereby making compliance monitoring more efficient. Targeted inspection programmes help deal with specific risks that are already identified, but they provide biased estimates of general compliance behaviour.

28. Sacrificing targeted inspections in significant numbers to replace them with random ones (given the same amount of resources) can have an adverse impact on fulfilling the law enforcement mission. There is a consensus among environmental enforcement practitioners that inspections should be primarily used to achieve compliance rather than to measure compliance. Indeed, there is no evidence of any country developing statistically valid compliance rates through the use of random inspections or by combining targeted and random inspections. It is, however, possible to generate statistically valid compliance rates for a segment of the regulated community, e.g. for high-risk installations which are inspected most frequently. EEAs in many countries rely on self-reporting to gather environmental performance data across all entities subject to a particular regulatory regime. There is, however, a risk that relatively minor violations will be under-reported.

⁴ Close to the concept of compliance rates is that of recidivism – a percentage of a certain universe of past offenders which revert to illegal behaviour during a specified observation period. This measure has been used by the US EPA as a criminal enforcement performance measure since fiscal year 2010.

⁵ In theory, compliance rates can be statistically valid if data can be gathered from at least 80% of the population or if the compliance monitoring sample is representative of the regulated community (US EPA, 2006).

29. Despite the statistical bias of compliance rates and the difficulty of using them as a meaningful indicator of regulated entities' behaviour, there is still demand for them as a tool in programme targeting. Compliance rates (especially those calculated for specific segments of the regulated community or for specific regulatory requirements) can help programme managers to determine whether inspections strategies are succeeding in finding non-compliance, to develop and target inspection programmes, target outreach and technical assistance to the regulated community, target follow-up enforcement, and design industrial sector-specific initiatives.

30. Apart from a statistical bias, the binary nature of traditional compliance rates (one is either in or out of compliance) does not allow regulators to compare performance across regulated entities: if a facility has made changes to move towards compliance but has not reached full compliance, a binary compliance rate would not capture that progress. If a compliance rate treats compliance at facilities with different scales of environmental impact equally, an apparent high compliance rate can be misleading if the most significant pollution sources remain out of compliance. In order to address the issue of different severity of offences, it is necessary to diversify compliance indicators based on the degree of the seriousness of offences.

2.2 Categorisation of non-compliance

2.2.1 Degrees of environmental harm

31. EEAs in OECD countries are increasingly diversifying breaches (violations, contraventions) based on their relative significance. The obvious first factor taken into account is the degree of potential environmental harm.

32. The classifications vary in complexity. Environmental regulators in Denmark and Poland, for example, use only four categories of offences, from those with no direct impact on the environment to those resulting in serious pollution. The South Australia EPA distinguishes between, on the one hand, contraventions with a potential to harm the environment, human health or safety (with five levels, from minor nuisance to serious environmental harm) and, on the other, administrative non-compliance (e.g. breach of licence conditions not entailing environmental harm, late payment of fees, etc.). The degree of potential environmental impact is usually a matter of judgement by inspectors, reviewed by their peers.

33. Both the Irish EPA (Table 1) and the South Australia EPA categorise the potential environmental impact of a violation. In addition to qualitative criteria, South Australia uses the monetary value of actual or potential property damage resulting from the offence as an important criterion of its seriousness. Both agencies also account for the likelihood of impact, with the South Australia EPA linking it directly to the operator's compliance history.

Table 1. Severity of consequence of non-compliance: Classification by EPA Ireland

Level	Descriptor	Description
1	No impact	No impact on the environment
2	Minor impact	Short term or localised issue that does not cause environmental pollution, minor aesthetic issue or not health related. Minor effects on air quality, water quality, soil, groundwater, odour or noise.
3	Moderate impact	Moderate impact on the receiving environment. Deterioration of water/air/soil quality, odour or noise issues for a short duration (a number of days)
4	Significant impact	Significant impact on the receiving environment: <ul style="list-style-type: none"> • Fish kill • Deterioration of water/air/soil quality over an extended duration • On-going Odour/Noise pollution • Potential significant human health problems
5	Serious impact	Serious impact on receiving environment: <ul style="list-style-type: none"> • Breach of environmental standards • Serious air/water/soil pollution with long term effects • Public health problems • Serious reduction in amenity value

Source: EPA Ireland, 2013

34. While the Irish EPA and the South Australia EPA use non-compliance categories only to determine an appropriate enforcement response, the Environment Agency (England) has an explicit objective of diversified measurement of non-compliance. Its Compliance Classification Scheme (EA, 2013a) differentiates between four categories of permit breaches based on reasonably foreseeable potential environmental impact, from most serious to negligible.

35. The potential impact is assessed on a case-by-case basis, since a violation that could cause a significant impact in one location may have a lesser impact elsewhere, depending on the proximity and nature of sensitive receptors. The Environment Agency distinguishes 28 sub-criteria that cover the possible types of permit breaches. Where multiple permit conditions have been breached due to a single failure by the operator, these non-compliances are sometimes consolidated into one record.

36. It is worth noting that some EEAs, such as the Environment Agency and the Irish EPA, measure compliance only of permitted/licenced installations while others like the South Australia EPA, at least in principle, evaluate compliance with any applicable legal requirement.

2.2.2 Differentiation of licence conditions

37. A slightly different approach is distinguish permit/licence conditions governing direct impacts on the environment (i.e. pollution releases) from those dealing with the operator's environmental management practices. The Scottish EPA (SEPA) exemplifies this approach, where the severity of non-compliance is a function of the degree to which a certain licence condition was exceeded, while the actual environmental impact (e.g. fish kill or substantial exceedance of an ambient environmental quality standard) is taken into account only for significant breaches.

38. SEPA's compliance assessment system differentiates between breaches of *environmental limit conditions* (ELCs) and *environmental management conditions* (EMCs). Violations of ELCs can be minor, repeated minor, gross and significant, depending on the degree of exceedance of the limit. SEPA has set generic rules for assessing gross and significant breaches as well as repetitions of a minor breach, although

some regulatory regimes have different or specific additional rules due to fundamental differences in the type of activity or license requirements (SEPA, 2012).

39. EMCs are specific to each regulatory regime but fall under three main categories: management, infrastructure, and reporting and record-keeping. Environmental management conditions are not necessarily administrative (paperwork-related) and can also include operating plans, monitoring procedures and training requirements. Breaches of EMCs can be minor or major. At the same time, failure to submit a correct description of the installation, its activities or waste types (as related to the scope of the licence) is considered to be a gross breach of environmental limit conditions.

40. SEPA also determines overall compliance with regard to environmental management. A major non-compliance with any EMC or minor non-compliance with more than half of EMCs are interpreted as low performance. High-performance installations should have no major EMC breaches and only one or two minor ones, depending on the total number of environmental management conditions. All installations that do not fall into the low or high performance category are characterised as having medium performance.

41. One of the drawbacks of the Scottish scheme is the difficulty of defining licence conditions as those related to environmental limits or environmental management. Tying violations to particular licence conditions and determining the level of each breach can also be quite challenging for staff. A nationally agreed procedure complemented by training and senior management oversight help improve the consistency of the scheme's implementation.

2.3 Measuring overall compliance performance of regulated entities

42. The next step from measuring the degree of non-compliance is to measure the performance of individual regulated entities through some kind of composite rating which may combine quantitative and qualitative indicators (see also Section 3.2.2). This compliance rating may or may not be integrated with the more general assessment of environmental risk of individual regulated entities.⁶

2.3.1 Qualitative compliance rating

43. In Scotland, SEPA has established a stand-alone Compliance Matrix which associates compliance with environmental limit and management-related licence conditions and comprises six performance bands: excellent, good, broadly compliant, at risk, poor and very poor (Table 2). The scheme is applied by inspectors throughout the calendar year to compile each installation's compliance record, which is used the following year to determine the annual overall compliance band. Certain licensed activities are considered to be low-risk and are not routinely inspected each year. The compliance band allocated to such activities is carried forward for the years in which no inspections are conducted. However, the assessment can be revised on receipt of new information or reported data. SEPA is considering changing the calendar year assessment period to a rolling 12-month one to allow for continuous assessment of the compliance record.

⁶ This approach may not be applicable in situations where a regulated community is unknown or difficult to identify.

Table 2. Scottish EPA's compliance matrix

		Environmental Limit Conditions		
		No breaches	Minor breaches or one gross breach	Significant breach, more than one gross breach or repeated minor breach
Environmental Management Conditions	High performance	Excellent	Good	Poor
	Medium performance	Good	Broadly compliant	Poor
	Low performance	At risk	Poor	Very poor

Source: SEPA, 2012

44. The annual compliance assessments are published on SEPA's website and are searchable by activity sector, geographic area and regulatory regime. The compliance rankings usually get attention from the press, which creates public pressure on operators to improve their environmental performance.

45. A Scottish-type qualitative compliance assessment scheme is fairly easy to understand and to explain to the public, and serves its purpose as an important part of the agency's performance measurement system. The compliance assessment also contributes to site-specific risk assessment as part of the Dynamic Regulatory Effort Assessment Model (DREAM), which is used to determine planned baseline inspection frequencies and increased monitoring in response to poor performance.

2.3.2 Linkages between compliance assessment and risk assessment

46. In many EEAs, a substantial amount of compliance information is often collected for risk assessment and inspection targeting purposes but is not fed into the agency's performance measurement system.

47. For example, the Irish EPA developed an Environment-Based Assessment Tool to help it prioritise its enforcement activities. The methodology allocates an enforcement category to licensed installations on the basis of five environment-based attributes: complexity, emissions, location, operator management, and enforcement record. Three of these attributes are related to compliance: emissions (installations with more than three violations of emission/effluent limit values are assigned to the high enforcement category), operator management (accounts for the number of notifiable incidents) and enforcement record (Box 1).

Box 1. Scoring facilities' enforcement record in Ireland

The score received by a facility for its enforcement record depends on the following:

- Number of complaints received by the EPA;
- Number of non-compliances notified by the EPA. This refers to the number of licence conditions not complied with – the total is summed from each site inspection/audit report or other EPA notification;
- Number of section notices (compliance orders); and
- Whether or not there are soil or groundwater contamination issues on the site.

In scoring facilities on these factors, section notices and convictions are given a higher weighting than non-compliances, which in turn are weighted more highly than complaints. The enforcement category (high, medium or low) is assigned based on the total number of points. If a facility has been convicted during the 12 month period for which the assessment applies, the assessment outcome will be one category higher than otherwise obtained.

Inspection is not required for enforcement categorisation based on the attributes used within the methodology. Most of the required data are provided by the operator and are verified by inspection, either already undertaken or to be undertaken at a future date. In addition, any significant changes to enforcement categories for facilities are reviewed by the relevant inspector for that facility. All licence holders are informed of their overall enforcement category, and detailed assessment information is available to the public upon request.

Source : EPA Ireland, 2010

48. The Irish scheme has some elements of performance measurement, although the scoring mixes output indicators (number of enforcement actions) with outcome ones (number of non-compliances, complaints, etc.) and is not associated with a specific performance target. Moreover, the number of people complaining to an EEA about industry-related environmental incidents is not necessarily indicative of their number or seriousness and may have more to do with the location of industrial facilities in densely populated areas or the environmental engagement of local communities. The Irish EPA is currently developing and implementing a new Licensing, Enforcement, Monitoring and Assessment system (LEMA), which will allow automated compliance assessment – the entry of compliance data and adjustment of compliance attributes in real time – to be more closely integrated into performance measurement.

49. Integrating compliance measurement with risk assessment provides mutual benefits for the two management processes. The best example of this is how the Environment Agency (England) links its Compliance Classification System to the Operational Risk Appraisal (Opra) scheme through the compliance rating attribute (Box 2). The installation's compliance rating provides an overall compliance profile of an individual installation/site and affects its overall Opra risk score, which in turn is a major factor in targeting the Agency's compliance monitoring efforts and in assessing the annual permitting charge imposed on the operator.

50. The scheme is used directly by the Environment Agency's officers to assess, record and monitor compliance. Monitoring and reporting is carried out at both local and national levels. The difficulty arises in the consistency of use and identification and recording of compliance breaches on the system. The Environment Agency reports that increased training, guidance and support has improved consistency over the last few years.

Box 2. Compliance rating in the Opra scheme in England

Compliance rating identifies the risk arising from operating regulated installations/sites. The Environment Agency has adopted a standard approach to classify permit breaches known as the Compliance Classification Scheme (CCS). Compliance rating is based on the CCS events over the course of the previous calendar year (January to December).

For each permit breach recorded in the CCS under one of the four categories based on potential environmental impact (from Category 1 corresponding to most serious offences), a certain number of points is allocated as follows:

CCS category CCS Points per breach

1	60
2	31
3	4
4	0.1

The points from each event are added up to produce an annual total non-compliance score, which is then converted into a compliance rating (an Opra band) as follows:

Band A: 0 CCS points.	Band D: 30.1 to 60 CCS points.
Band B: 0.1 to 10 CCS points.	Band E: 60.1 to 149.9 CCS points.
Band C: 10.1 to 30 CCS points.	Band F: 150 CCS points and above.

The compliance rating attribute affects the installation's overall Opra risk score. The risk score is reduced if the installation received an A rating. This reflects the reduction in risk posed by well-managed installations. An occasional minor breach of a permit condition (B rating) does not change the installation's Opra risk score. For installations having a band higher than B, the Opra risk score increases.

The system provides a clear focus for the Environment Agency's regulatory effort by identifying the worst performing sites: those in bands D, E and F. The system is reactive, and sites can move up and down the performance bands. The Agency has a corporate scorecard target to reduce the number of D, E and F performing sites.

Individual operators' compliance status is published on the Agency's website using "What's in your backyard" interactive maps. The Agency's Sustainable Business reports contain sector compliance performance data.

Source : EA, 2014: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/301770/LIT_6665.pdf

51. In addition to linking the compliance rating to each permitted installation's risk score, the Environment Agency considers the operator's compliance history as part of Opra's operator performance attribute. It uses a rather complex algorithm to calculate "penalty points" depending on the type of environmental enforcement action taken in each case of non-compliance.

52. Despite its methodological robustness, the complex design of the Environment Agency's compliance and risk assessment schemes requires a high level of data quality assurance and relatively sophisticated data management, which may prove excessively costly for smaller EEAs. The challenge of reducing the data management burden is addressed in Section 5.2.

2.4 Possible compliance targets and indicators

53. The differentiation of compliance "marks" allows an agency to broaden the menu of its compliance indicators. The traditional (binary) measure of the percentage of facilities (sites, installations) identified as non-compliant – still necessary to measure the appearance of new non-compliant sites – can

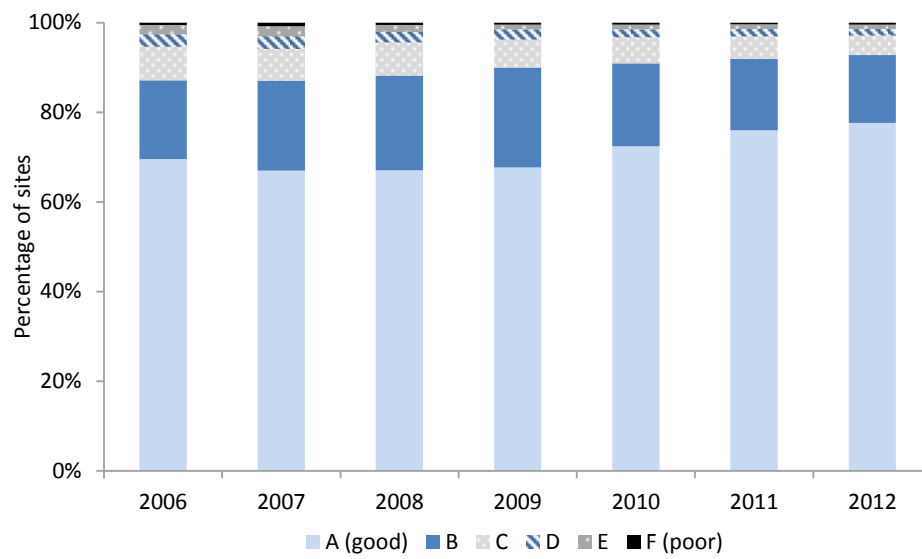
be complemented by other, more meaningful indicators. For example, it is possible to measure the occurrence of more or less serious non-compliance. The Scottish EPA goes even further by setting targets for and measuring the improvement, from one year to the next, of performance at non-compliant sites (e.g. improvement at 50% of sites identified as non-compliant in 2012).

54. The Environment Agency (England) sets targets for the number and percentage of sites to be compliant across a range of activities and industry sectors (e.g. 5% reduction of the number of breaches per year). These are aggregated for corporate reporting as well as reported and analysed separately by industry sector. The indicator “More businesses comply with permit conditions” measures the number of Category 1 and 2 (most serious) breaches, according to the Compliance Classification Scheme, and covers breaches of environmental permits, wastewater discharge consents, water abstraction licences, etc., reflecting the principal regulatory regimes.

55. The number of serious and significant pollution incidents (categories 1 and 2 according to the Agency classification) is used in England as a surrogate measure of environmental impact. Another advantage of this indicator is that it covers *all* sites, whether or not they fall under the Environment Agency’s permitting or licensing regimes. For example, there were 504 serious and significant pollution incidents in England in 2012, down by 8% compared to 2011 (thereby exceeding the 5% annual reduction target). The occurrence of pollution incidents is then further broken down by sector (farming, water companies, waste management facilities, other industry, etc.), between permitted and non-permitted sites, and between serious and significant incidents (EA, 2013b). An interesting supplementary measure is the number of serious pollution incidents per 100 permits in a sector, which shows the proportion of incidents in each sector and highlights sectors which cause a disproportionate number of incidents. However, this indicator is heavily dependent on the compliance of regulated entities with an obligation to report significant pollution incidents to competent authorities, since not all incidents can be detected by the EEA itself or through complaints by third parties.

56. In addition to these single-measure indicators, a compliance rating scheme with several performance “bands” offers the EEA an opportunity to measure and report on the share of regulated entities in each band and their movement across the bands over time. For example, the Environment Agency reported in 2013 that 78% of the sites it permitted in 2012 were rated A, compared to 76% in 2011 and 70% in 2006 (EA, 2013b). A dynamic evaluation of the regulated community’s compliance behaviour sends an even more powerful message about the outcomes of the agency’s work (Figure 1).

Figure 1. Compliance with permits in England by performance band, 2006-2012



Source: EA, 2013b

3. COMPOSITE INDICATORS

3.1 Advantages and disadvantages of aggregate measures

57. By combining the information contained in several indicators, composite indicators (otherwise referred to as aggregated indices) make it possible to convey simple messages about complex phenomena. Among their strengths is the potential to simplify the public communication process and allow policy makers to get highly condensed information on a multi-dimensional issue (such as environmental compliance) without getting lost in detail. Compliance indexing approaches can capture changes in performance that are more nuanced than a switch from compliance to non-compliance (or vice versa) reflected in a binary compliance rate. An index or a score can describe the extent of compliance by measuring it on a scale and allows regulators to compare performance across facilities.

58. However, most EEAs in OECD countries do not use composite indicators. The main reasons for this are the following:

- A large volume of information tends to be collected (at a significant cost) but then never “reaches the surface” to be analysed and discussed, so real problems requiring action may go unnoticed;
- Reducing the number of indicators by condensing information runs the risk of misinterpretation because users are not always aware of the scope and limitations of the index methodology, and because the message conveyed may be distorted by data gaps;
- Composite indicators make are difficult to integrate into the organisation’s strategic planning, either by setting appropriate overall targets or differentiated ones for the component indicators;
- An index using a complex weighting system (so that all component indicators are not treated as being of equal importance) becomes a mathematical exercise rather than a clear, actionable statement of performance.

59. Still, despite the costs and risks associated with composite indicators, they continue to be developed by environmental authorities in several countries mostly due to their attractiveness as a communication tool targeting external stakeholders. The mixed experience of their use to-date is described in the following section.

3.2 Experience with compliance-related indices

60. From the limited practical experience, it is possible to distinguish several categories of compliance-related composite indicators: pollution release indicators linking pollution load reductions to compliance assurance activities, facility-specific compliance indicators, and agency-wide performance indices.

3.2.1 Pollution release indicators

61. Reductions of pollution releases of regulated substances can be regarded as an intermediate outcome of compliance assurance activities, especially if those substances represent explicit priorities for

the agency. There is some academic literature on designing an Environmental Emissions Index (Styles et al, 2009), but very few practical applications.

62. One example is the Pollutant Load Indicator used in New South Wales, Australia – a total pollutant load of 12 air and 17 water pollutants emitted for the reporting year by all facilities required to pay load fees under the Load Based Licensing Scheme, weighted to reflect the relative harm of the pollutants and the sensitivity of the environment (“critical zones”) into which they are emitted. However, the Pollutant Load Indicator is not directly linked to enforcement actions (Mazur, 2010).

63. Environment Canada introduced in 2010 an Enforcement Environmental Improvement Index that was supposed to measure the mass of regulated substances reduced through enforcement actions, measured in equivalent metric tons of reduced substance. It was expected to gradually integrate the releases of over 40 air and water pollutants weighted in accordance with their toxicological impact (using the inverse values of respective ambient environmental quality standards as coefficients). Specific coefficients were also set for global-impact pollutants (such as greenhouse gases and ozone-depleting substances) that do not have toxicity-related standards (Mazur, 2010). Essentially, the scheme would first account for reductions of individual regulated pollutants as a result of compliance assurance activities and then try to aggregate them into a composite measure characterising the environmental impact of these reductions. Data had to be obtained during enforcement activities and recorded in the Environment Canada enforcement database. However, Environment Canada found it difficult to argue convincingly for a causal link between enforcement activities and environmental outcomes, particularly in trying to make a claim about preventing potential or real environmental harm via some enforcement action. Overall, this complex approach proved excessively costly and was abandoned shortly after its introduction.

64. This experience seems to suggest that focusing on priority pollutants rather than aggregating the data for a wide range of parameters and/or pre-weighting of pollutants via pollutant characterisation methodologies can make the results more meaningful in illustrating progress in achieving EEA goals.

3.2.2 Facility-specific compliance indicators

65. An attempt to introduce weighted compliance rates was made in 2008 by the Ministry of the Environment of Canada’s Ontario province. Its Compliance Index was a weighted sum of individual facilities’ violations of legislative provisions and permit (“certificate of approval”) conditions, with weights assigned to violations of each of approximately 1,300 legislative provisions. Each violation was supposed to be classified as one of four “contravention categories” (reporting and recordkeeping, operating standards, monitoring and sampling, or exceedance of emission/discharge limits) and assigned a corresponding weight. These weights would distinguish between procedural and substantive non-compliance, thereby reflecting to some extent the level of potential environmental impact from the offence (higher weights represent a larger impact). However, this compliance index has not been implemented, partly because of its complexity and partly due to the need for an expensive random sample approach to inspection planning (IEC, 2010).

66. Compliance performance scores and bands implemented in the UK and Ireland and described in Section 2.3 can also be considered composite indicators. Unlike Ontario’s failed compliance index, the English and Scottish schemes have a stronger qualitative element, which makes them more user-friendly while at the same time leaving space for inspectors’ judgement and raising the risk of inconsistency. The administrative costs of such schemes (including the costs of staff training) remain a concern, although these costs are lower if the schemes build on the existing operational practice, for instance, of assessing the seriousness of non-compliance.

67. Risk scoring, following the example of the Environment Agency's OPRA scheme, is spreading across OECD countries, mainly because of the resource-driven need for risk-based inspection targeting. Although compliance is usually only part of a risk score, such scores can be justifiably used as outcome indicators of compliance assurance. The dynamic (trends) analysis of movements between compliance performance and risk scores/bands enhances considerably their policy relevance.

3.2.3 Agency-wide performance indices

68. While the author is not aware of any indices intended to characterise the performance of the entire EEA in OECD countries, the recent experience of China in testing such an index is instructive. China's Ministry of Environmental Protection (MEP) has been conducting a pilot programme on performance evaluation of environmental enforcement, where provincial and local Environmental Protection Bureaus would use a 100-point reference performance index combining 72 input, output, intermediate outcome and final outcome indicators. Annex 1 contains more details about this pilot programme.

69. The pilot implementation of the index has faced many challenges, including deficiencies in the design, low data availability and credibility, and the lack of comparability across enforcement authorities, all of which are discussed in Annex 1. Most importantly, however, the MEP is realising that comparing aggregate scores of different agencies is next to impossible, while the index conceals a lot of meaningful information. As the MEP is looking to refine the design for further testing, it is pondering the possibility of disaggregating the index into several components that would better correlate, on the one hand, input and output indicators, and on the other, input and outcome indicators. Another lesson learned from this experience for China is the need to identify performance measures that should be considered individually, with an emphasis on intermediate outcome indicators.

70. More generally, composite measures seem to be analytically sound only when they aggregate indicators of the same type (e.g. intermediate outcome measures) describing the same phenomenon (e.g. compliance with permit requirements). The misperceived benefit of having one number to serve as a "mark" of the agency's performance is likely to be outweighed by the costs of collecting and processing large volumes of information that may not be relevant for performance management purposes.

4. CROSS-JURISDICTIONAL COMPARABILITY OF OUTCOME INDICATORS OF COMPLIANCE ASSURANCE

4.1 Drivers and constraints of comparability and benchmarking

71. In many OECD countries with decentralised environmental governance⁷, the national environmental authority has a mandate to oversee the implementation of national environmental legislation. This is usually done by reviewing the performance of sub-national EEAs (in a three-tier institutional structure, sub-national authorities should do the same with respect to municipal agencies).

72. However, national environmental authorities generally have very little quantitative information on the lower level EEAs' activities and results. The lack of routine compliance assurance data reporting (apart from annual reports produced primarily for public relations purposes), most often due to the political and/or institutional autonomy of sub-national and local competent authorities, is one of the key obstacles to the establishment of systematic external EEA performance reviews. Another impediment is the dissimilarity of performance indicators that sub-national and local EEAs use for their internal performance management (which would otherwise have been a good entry point for interagency comparison).

73. An analysis of strategies and instruments for promoting consistency in the implementation of national environmental law in decentralised systems (Mazur, 2011) has revealed an important lack of sound outcome indicators (and guidance for their interpretation) that would constitute a basis of comparison across jurisdictions. Data collected by sub-national EEAs, whether reported to the national government or not, is almost always focused exclusively on inputs (the number of inspectors and the amounts of allocated resources) or outputs (activities such as the number of inspections or enforcement actions taken). Whereas many sub-national and local EEAs argue that it is not the means of achieving compliance that necessarily need to be consistent but the results of their activities, very few use outcome indicators. Developing a limited set of comparable outcome indicators as part of interagency collaboration and adding them to the scope of "vertical" performance reviews would require significant resources from those agencies.

74. While benchmarking across jurisdictions may allow national authorities, the regulated community and the public to know whether there is equality in regulatory and compliance assurance activity, it is often resisted by sub-national and local EEAs. Many regional and local environmental regulators resent being seen in an unfavourable light based on what they usually claim to be inaccurate or poorly interpreted data. This challenge is even greater internationally, as in the case of establishing parameters for reporting by EU member states to the European Commission on the effectiveness of national environmental compliance assurance systems.

75. In order to address these barriers, it is essential that a selected set of comparable outcome indicators be quite small and supported by adequate performance measurement and data collection and reporting mechanisms as well as an interagency consensus for its implementation. National and

⁷ Decentralisation of environmental governance may mean delegation of responsibilities for the implementation of national legislation to the regional or local level and/or compliance assurance with respect to sub-national and local regulations.

international networks of environmental regulators can offer a vehicle for reaching such consensus through leadership and coordination.

76. Such a limited set of comparable indicators may also be developed bilaterally by two sub-national EEAs wishing to peer-review their respective practices and to learn from each other. This kind of bilateral benchmarking can be later joined by other interested authorities, thereby representing a bottom-up approach to building comparability into outcome performance measurement.

4.2 Potentially comparable outcome measures

77. Many compliance indicators discussed earlier in this paper (e.g. in Section 2.4) can be used to compare performance across EEAs. It may also be possible to conduct inter-agency comparison with the help of certain indicators of pollution releases or improved environmental management practices. However, there are different degrees of practical limitations associated with each type of measures.

4.2.1 Measures of serious non-compliance

78. Once the EEA has categorised non-compliance and defined serious (significant, major) violations, it is possible to measure the percentage of inspected facilities (installations, sites) found in serious non-compliance. Ideally, the definition of serious offences would be the same across jurisdictions, but in the absence of such common definition the comparison could be made by looking at the trend of the percentage (rather than the absolute number) of serious non-compliance across the regulated community.

79. It is also important to account for the differences in the definition of the regulated community. One way of doing this is to limit the scope of comparable indicators to permitted or licensed facilities (installations, sites), since it can be quite difficult to find a comparable way of defining the regulated community beyond the application of permitting or licensing requirements for key regulatory regimes. This, however, would not prevent individual EEAs from measuring serious non-compliance among non-licensed facilities, particularly those that are subject to systematic, sector-specific inspection campaigns.

80. Compliance indicators could be further refined by focusing on a limited number of “marker” requirements of the national legislation – priority norms that are the same in every sub-national jurisdiction. Internationally, such “marker” requirements would correspond to specific enforceable duties under key environmental Directives of the European Union.

81. Composite measures of compliance (scores and bands discussed in Section 3.2.2) can be powerful indicators of the effectiveness of regulatory effort. Indicators representing the share of regulated entities in each compliance band (with a breakdown by activity sector) and especially their movement across the bands over time can be comparable despite the potential differences in compliance categories across jurisdictions. At the same time, a significantly different number of categories defined by different EEAs would impede the comparison.

82. As with any compliance rates, the two principal limitations of these measures are the statistical bias due to the common practice of targeting inspections on facilities likely to present serious compliance problems (discussed in Section 2.1) and the ambiguity of interpreting a low or falling rate of serious non-compliance as successful prevention or poor identification of offences.

4.2.2 Occurrence of significant pollution incidents

83. The number of pollution incidents can be measured in relation to the number of permitted facilities with a further breakdown by sector and incident severity. This measure allows benchmarking of

the EEAs' effectiveness in preventing serious incidents. Similarly to indicators of serious non-compliance, the potential differences in definitions of severity could be ignored in the context of trend analysis.

84. However, a significant number of serious pollution incidents (around 40% in England) come from non-permitted sites. To make an inter-jurisdictional comparison valid, those facilities would either have to be defined similarly (the feasibility of this would depend on the degree of regulatory discretion of sub-national environmental authorities) or excluded from the comparable indicator. As noted in Section 2.4, it is entirely possible to measure the occurrence of serious pollution incidents in absolute values among permitted and non-permitted facilities within each jurisdiction.

4.2.3 Pollution release indicators

85. The United States Environmental Protection Agency (US EPA) measures the total mass (million pounds) of air pollutants (or water pollutants, or toxics and pesticides, or hazardous waste) reduced, treated, or eliminated through concluded enforcement actions. These indicators describe environmental performance improvements likely to translate into improved environmental conditions. However, they represent aggregations of large numbers of different pollutants with various toxicity, exposure and absorption characteristics.

86. Focusing on priority pollutants (for example, England's Environment Agency measures reduction of emissions of eight priority air pollutants) rather than aggregating the data for a wide range of parameters makes the indicators more meaningful and better attributable to the EEA's activities. It is conceivable that certain air and water pollutants or toxic substances are declared to be national priorities, in which case sub-national EEAs can be compared using trends of releases of these substances. More often, though, different EEAs will have their own priorities and resource allocations, making the cause-and-effect relationships between their activities and releases of "national priority" pollutants rather tenuous.

4.3.2 Monetary value of complying actions

87. Another high-profile indicator used by the US EPA in its performance assessment programme is monetary value of complying actions of the regulated community. The "dollars invested in improved environmental performance or improved environmental management practices as a result of concluded enforcement actions" indicator is unique in trying to express environmental outcomes of enforcement in monetary terms. This measure covers investments in technology improvements, better management practices, and adding staff and/or hiring contractors for environmental purposes. Although a financial investment does not necessarily translate into improved facility performance, the "dollars invested" indicator is a legitimate measure of the regulated entities' commitment to improve their environmental performance, with a potential to prevent future environmental harm.

88. Theoretically, monetary measures should provide a good basis for inter-jurisdictional comparison. However, there is currently no evidence of any EEA outside the United States collecting information from regulated entities regarding the monetary value of their corrective actions. In some countries such information may be considered as commercially sensitive and, therefore, difficult for an EEA to obtain. Furthermore, a high value of this indicator may be interpreted as a reflection of high compliance costs of the regulated community. Getting credit for the EEA for imposing an allegedly excessive compliance burden on businesses is usually politically difficult, especially since the most expensive compliance action may not always deliver the best environmental outcome. In addition, it is impossible to set a target related to such an indicator, as the monetary value of complying actions is case-specific and hardly predictable.

5. OPTIMISATION OF A SUITE OF INDICATORS

5.1 Reducing the data management burden

89. Many EEAs starting to develop a system of performance indicators tend to try to formulate and track a large number of measures that would reflect almost every aspect of their activities, using the maxim “You can’t manage what you don’t measure”. As the performance measurement system becomes institutionalised, it often appears that it is overly burdensome for staff in terms of time and resources – data entry can be tedious and is often considered less important than doing the compliance and enforcement work that generates the data. Moreover, many of the originally designed indicators often appear not to be crucial for the agency’s management process. The constant pressure to “do more with less” and reduce operating costs contributes to the need to re-evaluate the suite of performance indicators.

90. For example, the Environment Agency (England) has in recent years sharply reduced the number of its internal management (“second tier”) indicators – from about 600 in 2009 to less than 100 in 2014. Similarly, the number of measures in the agency’s Corporate Scorecard (addressed to external stakeholders) went down from 58 to 21 during the same period. The streamlining and trimming of the indicator suite included the following key steps:

- Eliminating redundant indicators and retaining only the measures that are truly useful in the agency’s performance management process;
- Focusing on measuring outcomes and demonstrating their relationship with the agency’s activities (outputs);
- Centralising performance measurement by concentrating the processing and analysis of data in the head office while at the same time reducing the reporting burden for the field staff; and
- Maximising IT support for data management.

91. In another example, EPA Victoria (Australia) has faced a situation quite typical for an early stage of a performance measurement programme. It has put in place a sophisticated performance data management system (Box 3) and engaged in extensive efforts to expand its staff’s capacity. However, three years after the introduction of a first set of about 160 indicators⁸ the agency was facing an explosion of data collection, with many field staff wondering whether “it was all about the numbers” and not seeing the link between their daily duties and the corporate performance measurement. EPA Victoria is also launching a review process to re-evaluate the reasons for individual indicators to be tracked with an aim of making the performance measurement system leaner and more meaningful. One way to do this would be follow the example of the South Australia EPA which uses a Program Logic Model to link outcome measures with higher-level objectives.

⁸ EPA Victoria has less than 400 staff and significantly less resources than England’s Environment Agency with over 11 000 staff.

Box 3. Performance data management at EPA Victoria

EPA Victoria has established clear operating procedures for its performance measures and provides guidance to staff involved in collecting and collating data. Clear controls exist around data input procedures to facilitate consistency, and an extensive training programme has been implemented to assist staff using the system.

The datasets that EPA uses are documented in its indicator detail sheets and scorecards. The sheets identify the details of the data source used for each performance measure and provide instructions on how data is to be collected for each. The Integrated Business Information System (IBIS) provides EPA with a high level of confidence about the accuracy of collected and reported data. IBIS has an automated centralised data depository.

EPA has also nominated data stewards who are appointed for each data set and a scorecard owner for each performance indicator and output measure. The data stewards undertake scheduled reviews of the veracity of data content. They are responsible for overseeing the data collection processes. The stewards group meets regularly and reports to a data owners committee.

Source : Victoria, 2013

92. Irrespectively of the size of the indicator suite, compliance assurance outcome indicators, as any other performance measures, need to be regularly reviewed and revised to maintain their objectivity (as agency staff tend to adjust their work to maximise “positive” indicators) and relevance to the changing regulatory programme.

5.2 Adapting indicators to different audiences

93. Within a performance measurement framework, different sets of indicators are normally used for different purposes and audiences. Environmental compliance assurance indicators can be used by different categories of stakeholders:

- *Managers within the agency* can use indicators to assess their unit’s performance, adjust compliance assurance strategies to increase their impact, justify the need for, and optimise the use of, human and financial resources. Such internal planning and management indicators are sometimes referred to as a “dashboard” or “second-tier” indicators and usually require monthly reporting.
- *Policy makers* are particularly receptive to quantitative performance indicators which allow them to take measure of the effectiveness and efficiency of compliance assurance activities and substantiate their budget allocation decisions. These indicators generally constitute a corporate “scorecard” of “first-tier” indicators having primarily external accountability purposes and require less frequent, quarterly or biannual, reporting.
- *The public* primarily wants to know about results of compliance assurance activities, both in terms of behaviour of polluters and the improvement of environmental conditions. The regulated community typically constitutes a sub-set of this audience, but may be a target of specific output indicators demonstrating the EEA’s diligence in detecting and punishing non-compliance, which could help create a deterrent effect against future offences. Indicators targeting the public are commonly part of the corporate scorecard but may be published just on an annual basis.

94. Given the difference of objectives and target audiences, the EEA should review the “policy relevance” of individual measures to check whether they fit their purpose and audience, are easy to

interpret, and show trends over time. EPA Victoria, for example, uses social research to put its performance findings in a wider societal context and test their relevance. The usefulness of performance measures communicated to the public can also be assessed through opinion surveys (practised, for example, by local environmental authorities in Sweden) about the effectiveness of EEA's work.

5.3 Challenges of quantifying costs and benefits of performance measurement

95. In optimising the suite of compliance assurance indicators, it would be ideal for the EEA to be able to weigh the costs of tracking a certain performance measure against the benefits it brings to the organisation. However, in practice such analysis is hardly possible. While a number of EEAs in OECD countries have launched "value for money" projects, they find it difficult to monetise even the benefits of their core activities, let alone those of performance measurement.

96. One early example is a current pilot study of value added of England's Environment Agency's activities using dependency modelling which aims at identifying plausible causal links between activities and outcomes. This study suggests that a 10% reduction of strategic oversight and planning efforts (which cost GBP 175,000 in 2012-2013) would reduce the probability of achieving the Agency's policy goals by 1% and result in a lost value of GBP 1.1 million (EA, 2014).

97. Assessing the costs of pursuing concrete indicators is even more challenging. There is usually no budget allocation specifically for the purposes of performance data acquisition and management. Rather, the information is routinely collected as part of the overall compliance assurance programme. Very few EEAs have task-specific staff time accounting. Those that do (like the Environment Agency) are able to quantify the amount of staff time (and, therefore, money) spent on planning and performance measurement, but without a breakdown by individual measure.

98. The measurability of performance indicators is a good way to qualitatively assess their cost to the agency. An earlier study (Mazur, 2010) has found that in many EEAs resource limitations for data collection and management present the greatest barrier in developing outcome indicators. Therefore, with limited (and often shrinking) budgets, EEAs would tend to focus on compliance assurance indicators with the highest measurability, i.e. those data for which are readily available, and those that are most closely linked to existing input and output indicators.

99. Pollution release and pollution incident indicators have the highest measurability as they are usually based on routine emission reporting and incident notification as well as on expected outcomes of enforcement-induced corrective actions (Mazur, 2010). Compliance measures discussed in Section 2.4 are normally derived from inspection data or operators' self-reporting. Such intermediate outcome indicators would likely be most cost-effective.

100. The measures of environmental quality improvements are based on regular environmental quality monitoring and can be powerful and cost-effective final outcome indicators, if the EEA is able to show that they correspond to targeted, problem-oriented compliance assurance efforts.

101. On the other hand, for example, measures of effectiveness of compliance assistance are very cost-intensive: in most cases, data collection requires user surveys or on-site observations. However, these costs may be warranted in order to justify funding for the agency's compliance promotion activities.⁹

⁹ There is growing evidence that compliance promotion activities, particularly those targeting small and medium-sized enterprises save money for both regulatory agencies and regulated entities (Eftec, 2008).

BIBLIOGRAPHY

- EA (2011), *Opra for EPR version 3.8*, Annex A – Opra Scheme for Installations April 2013/14, Environment Agency, Bristol, UK.
- EA (2013a), *Recording non-compliance using the Compliance Classification Scheme (CCS)*, Operational instruction 534_10, Environment Agency, 26 March 2013.
- EA (2013b), *Sustainable Business Report 2012*, Version 1, November 2013, Environment Agency, Bristol.
- EA (2014), *Value-added of EA activities: results of the waste pollution incidents pilot*, Presentation by Chris Durham, Principal Economist, Environment Agency, 28 January 2014.
- Eftec (2008), *Economic Evaluation of the Benefits of NetRegs to Small and Medium Enterprise Users in the UK: 2008 baseline and future valuation method*, Final Report, Eftec, London.
- EPA Ireland (2010), *Guidance on Completion of Methodology for Determining Enforcement Category of Licences*, Office of Environmental Enforcement, Environmental Protection Agency, Ireland, 16 June 2010.
- EPA Ireland (2013), *LEMA Licence Enforcement Compliance Investigation Procedure*, 02-CI-LEMA, Environmental Protection Agency, Ireland, 25 October 2013.
- IEC (2010), *Compliance Indexing Project*, report prepared for US EPA, Industrial Economics, Inc., Cambridge, Massachusetts.
- Mazur, E. (2010), “Outcome Performance Measures of Environmental Compliance Assurance: Current Practices, Constraints and Ways Forward”, *OECD Environment Working Papers*, No. 18, OECD Publishing, <http://dx.doi.org/10.1787/5kmd9j75cf44-en>
- Mazur, E. (2011), “Environmental Enforcement in Decentralised Governance Systems: Toward a Nationwide Level Playing Field”, *OECD Environment Working Papers*, No. 34, OECD Publishing, <http://dx.doi.org/10.1787/5kgb1m60qtq6-en>
- OECD (2002), *Aggregated Environmental Indices: Review of Aggregation Methodologies in Use*, [ENV/EPOC/SE\(2001\)2/FINAL](http://dx.doi.org/10.1787/5kmd9j75cf44-en), Organisation for Economic Co-operation and Development, Paris.
- SEPA (2012), *Compliance assessment scheme guidance manual*, Version v 3.0, Scottish Environment Protection Agency, January 2012.
- Styles, D. et al. (2009), “Measuring the Environmental Performance of IPPC Industry: Devising a Quantitative Science-based and Policy-weighted Environmental Emissions Index”, *Environmental Science and Policy*, 12 (2009), 226-242, <http://dx.doi.org/10.1016/j.envsci.2009.02.003>
- US EPA (2006), *Expanding the Use of Outcome Measurement for EPA’s Office of Enforcement and Compliance Assurance*, Report to OMB, US Environmental Protection Agency, Washington D.C.

ENV/EPOC/WPIEEP(2014)13/REV1

Victoria (2013), Environment and Sustainability Sector: Performance Reporting, Victorian Auditor General's Report, Victorian Government, June 2013.

ANNEX. PERFORMANCE EVALUATION OF ENVIRONMENTAL ENFORCEMENT IN CHINA: FIRST LESSONS LEARNED

1. Objectives of the pilot programme

102. China has been putting an emphasis on increasing the capacity of environmental enforcement agencies since their establishment more than two decades ago. This was especially true during the 11th Five-year Plan (2006-2010), when over 3.2 billion CNY was invested in equipping inspectors with vehicles and portable monitoring instruments, which has led to a sharp increase of inspection activities across the country. However, these efforts have not brought about a reduction of pollution levels, which is caused by large-scale urbanisation and the high proportion of heavy industry in the economy.

103. China's Ministry of Environmental Protection (MEP) recognises that sole reliance on financial injections is neither sustainable in the long term, nor will it lead to higher compliance of the regulated community. Several factors are limiting the efficiency and effectiveness of enforcement actions while consuming excessive resources: multiple redundant inspections of same facilities by enforcement agencies of different administrative levels with overlapping responsibilities, inconsistent inspection procedures generating useless information, etc.

104. To address these problems, the MEP announced in January 2012 a new initiative on enforcement capacity building focused on the effectiveness and efficiency of enforcement, rather than just financial resources. Performance evaluation was identified as an important part of this initiative as an instrument to diagnose implementation problems and provide feedback to policy makers.

105. In July 2012, the MEP's Bureau of Environmental Supervision sent to the provincial Environmental Protection Bureaus (EPBs) a "Notice on strengthening the effectiveness and efficiency of environmental enforcement and implementing a pilot on performance evaluation" which stated that the pilot programme was aimed at testing several aspects of a future nationwide enforcement performance evaluation system:

- Identifying an analytically sound indicator system;
- Establishing a link between performance evaluation and capacity building; and
- Making performance evaluation an instrument to promote accountability and effective and innovative management of enforcement agencies.

2. Pilot programme participants and key milestones

106. All provincial EPBs are required to participate in the pilot. Their responsibilities with regard to the pilot include:

- Developing implementation plans with specific tasks;
- Selecting, based on mutual agreement, at least two cities (which in turn had to select each two counties or districts) for participation in the pilot;

- Establishing a leadership and coordination team consisting of relevant department heads to manage the pilot;
- Adapting the reference performance indicators, their assessment criteria, weighting and data sources specified in the MEP's Notice to the local situation and capacity building needs (including removing and adding indicators);
- Providing training and assistance to the participating cities and counties; and
- Overseeing the programme implementation in the pilot jurisdictions.

107. Overall, a total of 116 city and county-level EPBs were selected to participate in the pilot programme.

108. The first phase of the pilot started in June 2012 and ended in December 2013. Participating city and county EPBs were expected to produce interim and final reports using the reporting template developed by the MEP, by February 2013 and February 2014, respectively. Based on these reports, provincial EPBs should draft a general review report. However, they are not required to evaluate their own enforcement performance.

3. Reference performance index

109. The reference indicator system designed by the MEP is a hundred-point index which consists of 72 indicators which fall into four categories: basic capacity (20 points), internal tools and procedures (20 points), administrative performance (outputs, 35 points), and effectiveness (results, 25 points).

- The *basic capacity indicators* characterise the resource inputs provided by local governments to the respective EPBs in terms of staff, budget, equipment and information systems. They are intended to evaluate the implementation of the central government's 2011 "capacity standard" for enforcement agencies.
- The *indicators on internal tools and procedures* mainly intend to measure the agency's management capability, which is supposed to include 15 rules: ten "traditional" ones (planning, reporting, record-keeping, etc.) and five "innovative" tools (e.g. a compliance classification system, recording of inspection and enforcement actions, standard enforcement procedures, compliance guidelines for key industries).
- *Administrative performance indicators* measure such inspection and enforcement outputs as the number of site visits conducted, pollution charges collected and fines imposed. This category also includes several qualitative measures (e.g. regarding the fulfilment of the inspection plan).
- *Enforcement effectiveness indicators* are supposed to measure intermediate and final outcomes. Intermediate outcome indicators cover behavioural changes of both enforcement officials and companies, while final outcome indicators describe the improvement of environmental quality and public satisfaction with the agency's activities.

110. This combination of indicators reflects different stakeholders' views on the concept of performance. From the perspective of superior agencies, performance measures should depict how and to what extent their orders are implemented by lower-level authorities. At the same time, EPBs at the local level tend to focus on specific outputs and intermediate and final outcomes. It should be recognised that the

pilot's performance indicators mark an important shift from China's traditional focus on measuring the implementation of "commands from above" to trying to characterise real results on the ground.

4. Reported implementation to-date

111. As the participation in the pilot at the local level is voluntary to a large degree, most pilot EPBs are actively engaged in coordinating its operation, data collection, training and outreach events, as well as improving their performance in line with the indicators.

112. While most pilot EPBs have adopted the MEP's reference indicator system unchanged, some have tailored it to their needs. For example, the Shandong Province EPB condensed the index to 56 indicators, arguing that other indicators go beyond provincial EPBs' responsibilities. In contrast, the Sichuan Province EPB added four new indicators to assess the implementation of local enforcement initiatives.

113. However, many pilot EPBs that did not modify the reference index have expressed their objections to some indicators, such as the number of administrative enforcement cases, amount of monetary penalty and revenue of pollution charges, insisting that their increase means only better detection and punishment of non-compliance rather than improved effectiveness of enforcement. Environmental quality indicators are also controversial as outcome measures, since they are affected by multiple factors that are beyond the enforcement agencies' control. At the same time, there are suggestions to include the number of trainings and not just the existence of a training programme at the agency.

114. The comparison of interim and final reports submitted by the pilot EPBs indicated that most of them are increasing their performance scores, particularly with respect to basic capacity and internal management. The capacity indicators reflect the growing willingness of local governments to allocate funding to EPBs, but not necessarily the enforcement agency's own performance. Regarding internal management, the growing numbers may be explained by some EPBs copying tools and procedures from peer agencies but do not necessarily mean their effective application.

115. The analysis of 29 complete pilot reports from seven provinces has revealed that outcome indicators always get high scores, which may not correspond to reality. For example, over 51% of all pilot EPBs have reported top scores for compliance of industrial facilities with environmental requirements. The absence of a mechanism to verify the data seriously undermines the credibility of such reports.

5. Key implementation challenges

Deficiencies in the design of the performance index

116. The performance evaluation index covers too many indicators which are often linked to conflicting objectives. As a result, the overall index as a weighted sum of these indicators becomes meaningless. For example, adding up input and outcome indicators gives the same index value with higher inputs and lower outcomes (indicating reduced effectiveness and efficiency) and with lower inputs and higher outcomes (pointing, on the contrary, to significant progress). The number of intermediate outcome indicators (measures of knowledge and behaviour of the regulated community) is insufficient, even though they have a closer causal link to compliance assurance outputs than environmental quality indicators. Some indicators are defined ambiguously: for example, the rate of recidivism is defined as the "number of enterprises having violated the same law at least twice a year", but it may refer to a particular requirement or any requirement contained in a certain law.

117. In addition, with a few exceptions, indicators of different policy relevance are assigned the same weight (one point each): the total weight of a category is almost always equal to the total number of

indicators in that category. This contradicts the weighting approach to the aggregation of indicators and reduces the policy relevance of this index even further.

Lack of comparability across regions

118. The comparability across sub-national jurisdictions should in principle help to identify best practices and encourage performance improvement through benchmarking. However, the discretion granted to the provinces, autonomous regions and municipalities to adjust the reference indicators, their weighting factors and scoring thresholds leads to incomparability between different pilot regions. The total points can only be comparable between different pilot EPBs in one region.

Low data availability and credibility

119. The MEP's specifications for the pilot programme provide only very general guidance on data sources for the indicators. Pilot EPBs have to decide themselves which data to use, and information collection methods are not consistent from one agency to another. In reality, many EPBs do not currently collect any information on a number of indicators, particularly intermediate outcome ones (e.g. the percentage of industrial pollution control facilities found in normal operation throughout the year). Even if there is no deliberate cheating, the ambiguous definitions of the indicators may lead to differences in the indicators' interpretation by different EPBs. The absence of mandatory verification of the data sources and accuracy by the higher-level authorities seriously undermines the credibility of performance measurement.

Low buy-in of the pilot enforcement authorities

120. The importance of performance evaluation for the overall improvement of the management of their organisations has not been fully appreciated by the leadership of the pilot EPBs. Instead, performance evaluation is often seen as a threat of sanctions, hence the desire to embellish the evaluation results, often by falsifying the underlying data. This fear of punishment stems from the long-standing tradition in China to reward or promote officials based on the completion of specific assignments. Another reason for the leadership's reticence is their unfamiliarity with the concept of performance measurement: a survey conducted among the EPBs' pilot project managers who participated in MEP-led trainings has unveiled that more than half of them had no prior experience or training in performance evaluation.

121. An additional challenge for local EPBs participating in the pilot programme is the lack of resources to dedicate to performance evaluation. The authorities are overburdened by their daily inspection and enforcement tasks and assign low priority to the collection and interpretation of performance-related data. lack knowledge and capacity to interpret the points and take use of the evaluation to manage performance.

6. Lessons learned

122. The pilot programme on enforcement performance evaluation bears many features specific to China's institutional traditions, some of which are now in the process of change. For example, it reflects the transition from directive-based and outcome-based management, which has caused some of the above-mentioned weaknesses in its design. However, despite the technical imperfections of the performance index, the pilot provides a vehicle for improved capacity building of local EPBs and more effective enforcement across the country. It also acts as a tool to support the implementation of new enforcement initiatives.

123. The pilot programme's experience points to a number of specific conclusions:

- *An incremental approach* is needed to build up the awareness of importance of performance measurement as well as the capacity to interpret its results and use them establish an effective performance management system in sub-national environmental enforcement authorities in China. This is a challenge in a unitary and centralised country like China, where the profound legacy of directives and task-based evaluation needs give way to result-oriented performance concept and methods. A gradual transition would also help address the data availability barriers, as many enforcement agencies have not even begun yet to establish such data collection systems required to support performance evaluation.
- A decentralised approach to performance-based management and result-oriented evaluation would give local EPBs the flexibility to be creative in choosing the means to achieve compliance results and to target their resources depending on the local environmental risks and compliance and enforcement priorities. Otherwise, enforcement agencies would seek to maximise their performance scores by focusing on “easy” compliance targets (e.g. by inspecting more minor pollution sources with fewer violations). Making performance evaluation an internal management process rather than a third-party assessment would also help to enhance the ownership of results by the leadership of individual EPBs.
- The pilot EPBs have recognised the value of intermediate outcome indicators as part of the performance index. Many provincial EPBs assigned higher weightings to those indicators. However, they need to be defined with more precision. For example, the indicator of recidivism was adopted by few pilot EPBs, the rest complaining about the lack of clarity in its definition.
- The current scoring system is complicated and differs across the provinces. The current attempt to bring the 72 indicators, which range from binary measures (yes/no) to ratios to absolute values, to a common denominator by using a scoring system has created great disparity between the pilot EPBs. As a result, comparing scores of different agencies is next to impossible, while the index conceals a lot of meaningful information. This problem can be addressed by reverting to a lower level of aggregation of performance information. This could be done, for example, by assessing each of the four categories of indicators separately, identifying performance measures that should be looked at individually, and better correlating, on the one hand, input and output indicators, and on the other, input and outcome indicators. In addition, it may not be feasible to retain environmental quality (final outcome) indicators as part of the performance evaluation system, as many enforcement agencies consider that they do not necessarily reflect the effectiveness of EPBs’ activities. Instead, the emphasis should be placed on intermediate outcome measures, such as indicators of compliance and pollution reduction as a direct outcome of enforcement actions.

7. Next steps in improving the performance measurement system

124. The MEP intends to further pursue the pilot programme. But before launching its second phase, the MEP would like to analyse and incorporate the lessons drawn from the first phase, both to improve the design of the system and raise local capacity to implement it. Despite the training sessions conducted at the outset of the pilot programme, pilot EPBs still have a long way to go before local enforcement officials understand the merits of performance evaluation, develop locally relevant indicators, learn how to interpret performance data to diagnose and address internal management problems. Therefore, in addition to further training, the next steps will include “learning by doing” and exchanging experiences among peer agencies.

125. In preparing the second phase of the pilot programme, the MEP intends to review and improve several aspects of the design and implementation of the indicator system, capacity building, and the use of evaluation results. For example, the MEP wants increase the role of intermediate outcome (compliance) indicators and have certain measures reflect the seriousness of violations (while gradually enhancing the

data availability). The roles of different levels of enforcement agencies would be adjusted, with higher-level EPBs playing a bigger role in establishing a verification mechanism to ensure the validity and reliability of performance data and acting upon evaluation results. At the same time, subordinated agencies would be responsible for the completeness and authenticity of performance reports. The number of indicators in the performance index is expected to be reduced in order to minimise the costs of data collection and encourage the local enforcement agencies to participate in the pilot. Beyond the reduced reference set of indicators, local EPBs will be expected to bear the costs of tracking additional performance indicators they would like to introduce. Similarly, if higher-level agencies want to evaluate the impact of a specific initiative, they should be prepared to pay the extra costs.