

Getting messages across using indicators

A handbook based on experiences from
assessing Sustainable Development Indicators

2014 edition

Getting messages across using indicators

A handbook based on experiences from
assessing Sustainable Development Indicators

2014 edition

*Europe Direct is a service to help you find answers
to your questions about the European Union.*

Freephone number (*):

00 800 6 7 8 9 10 11

(* The information given is free, as are most calls (though some operators,
phone boxes or hotels may charge you).

More information on the European Union is available on the Internet (<http://europa.eu>).

Luxembourg: Publications Office of the European Union, 2014

ISBN 978-92-79-27205-9

ISSN 2315-0815

doi:10.2785/36378

Cat. No: KS-GQ-12-001-EN-N

Theme: General and regional statistics

Collection: Manuals and guidelines

© European Union, 2014

Reproduction is authorised provided the source is acknowledged.

Foreword

‘An indicator is a summary measure related to a key issue or phenomenon and derived from a series of observed facts. Indicators can be used to reveal relative positions or show positive or negative change. Indicators are usually a direct input into Union and global policies. In strategic policy fields they are important for setting targets and monitoring their achievement.’⁽¹⁾ Indicators also help to present the most important and relevant features of a given issue or topic to policy makers, stakeholders and citizens. Current international initiatives such as the Post-2015 Development Agenda and in particular the definition of Sustainable Development Goals, confirm once more the growing importance of indicators in the light of ‘evidence based decision making’.

Indicators are essential components of the information infrastructure provided by **official statistics**. They are based on basic data or accounts and, like them, they are produced under strict quality assurance. Official statistics work with established common methodological frameworks, which ensure consistency over time and place.

Indicators as such contribute to reducing complexity and hence facilitate communication. Synthesizing their information by means of **indicator-based assessment methods** is one more step that greatly contributes to producing **clear messages** to policy makers, stakeholders and citizens. The use of easy-to-interpret assessment symbols is a key feature of such methods.

Indicator-based assessment methods are not only useful for the purpose of synthesizing information, they are also an essential tool when it comes to assessing progress towards **quantified targets**.

A close interaction between evidence and decisions include risks, which have to be addressed with quality assurance measures. For this purpose, the European Statistics Code of Practice and the UN Fundamental Principles of Official Statistics, in particular the principles of independence and transparency, play a crucial role.

The growing importance of indicators contrasts with a comparably small amount of methodologically oriented literature. This report comes therefore at the right time and it is urgently needed. It draws up a state of play of the major initiatives on indicator-based assessment in Europe. It should serve as a base for the way forward in this field, especially in the context of the SDGs, but potentially also for indicator systems in other domains of official statistics.

The work on indicator-based assessment developed and exposed in this handbook complements the methodological reflection on indicators currently carried out by Eurostat. This exercise will lead to a series of papers on statistical indicators.

Finally, I would like to address my thanks to the members of this Expert Group for the excellent and innovative job they have done. I would also like to address a special thank you to the Swiss Federal Statistical Office who took the initiative to create this Expert Group and who coordinated its work and hosted it in Neuchâtel.



Walter Radermacher
Director-General, Eurostat
Chief Statistician of the European Union

⁽¹⁾ European statistical programme 2013-17, Regulation (EU) No 99/2013 of 15 January 2013.

Acknowledgments

This document is the result of the work of the Expert Group on Indicator-based Assessment (EGIA), created in 2010 at the initiative of the Swiss Federal Statistical Office (FSO) and mandated by the Eurostat Working Group on Sustainable Development Indicators (WG SDI).

The overall objective of the EGIA is to foster developments in indicator-based assessment by gathering practitioners, allowing them to exchange experience and build on common knowledge. The Group based its work primarily on sustainable development indicators, but the outputs should be generally applicable to other areas.

The EGIA comprises a small number of practitioners in the field of sustainable development indicators from statistical offices, ministries and public organizations, who were invited by the FSO to participate on the basis of their experience in the field of indicator-based assessment and ability to contribute to a substantive part of the handbook and methodological developments. The costs were covered by the participants' organizations.

Editor:

Eurostat

Project coordinators:

Anne Boesch and André de Montmollin, Federal Statistical Office (FSO), Switzerland

Main author:

Anne Boesch, Federal Statistical Office (FSO), Switzerland

Authors:

André de Montmollin, Federal Statistical Office (FSO), Switzerland;

Anna Kulig, Statistics Netherlands;

Viveka Palm, Statistics Sweden;

Vincent Willi, Federal Statistical Office (FSO), Switzerland;

Natacha Zuinen, Federal Planning Bureau (until August 2012), Service Public de Wallonie (from September 2012), Belgium.

Contributors:

Viktoria Bolla, Eurostat;

Sarka Buyny, Federal Statistical Office (DESTATIS), Germany;

Stephen Hall, Department for Environment, Food and Rural Affairs (DEFRA), United Kingdom;

Regina Hoffmann-Müller, Federal Statistical Office (DESTATIS), Germany;

Florian Kohler, Federal Statistical Office (FSO), Switzerland;

Christa Mühlemann, Federal Statistical Office (FSO), Switzerland (until November 2012);

Vittorio Perduca, Eurostat;

Edward Tredger, Department for Environment, Food and Rural Affairs (DEFRA), United Kingdom (until September 2010).

Members of the Expert Group:

Belgium: Natacha Zuinen, Federal Planning Bureau (until August 2012), Service Public de Wallonie (from September 2012).
Eurostat: Viktoria Bolla (from September 2012), Graham Lock (until February 2012), Vincent Tronet (from March 2012 until August 2012).

France: Cécile Dormoy (until August 2011), Service de l'observation et des statistiques du Commissariat générale au développement durable.

Germany: Sarka Buyny (until September 2010), Regina Hoffmann-Müller, Kerstin Manzel (from October 2010 until June 2012), all Federal Statistical Office (DESTATIS); Frank Hönerbach, Federal Environment Ministry.

Netherlands: Anna Kulig, Statistics Netherlands.

Norway: Tone Smith (until July 2010), Per Arild Garnåsjordet (from August 2010 until July 2012).

Sweden: Viveka Palm, Statistics Sweden.

Switzerland: Anne Boesch, André de Montmollin, Vincent Willi, all Federal Statistical Office (FSO).

United Kingdom: Stephen Hall, Department for Environment, Food and Rural Affairs (DEFRA).

United Nations Economic Commission for Europe (UNECE): Tiina Luige (from September 2012).

Layout:

Imprimerie centrale, Luxembourg

Language editing:

Kimberly Way, Toronto, Canada

Proofreading:

Jana Wachtl, Wachtl IND+, Binningen, Switzerland

Contents

Foreword	3
Acknowledgments	4
Introduction	8
Context of this handbook	8
Aim and structure of this handbook	8
1. The role of indicators and assessment	11
1.1 Policy context	12
1.2 The need to provide a summary or a comprehensive assessment	13
1.3 Contribution of indicator-based assessment to policy evaluation	14
1.4 Role of official statistics	14
2. Methodology of indicator-based assessment	15
2.1 Terminology	16
2.1.1 Terms terminology	16
2.1.2 Charts terminology	18
2.2 Method classification	18
2.2.1 Category 1: target value and target year available	20
2.2.1.1 Target path	20
2.2.1.1.1 Linear target path	21
2.2.1.1.2 Non-linear target path	22
2.2.1.2 Statistical projection: Extrapolation	23
2.2.1.3 Methods comparison	24
2.2.2 Category 2: target value available, no target year defined	27
2.2.2.1 Methods comparison	28
2.2.3 Category 3: no target value defined, rate of change available	29
2.2.4 Category 4: no target value defined, direction available (trend assessment)	30
2.2.4.1 Calculation of the observed evolution with initial and final values	31
2.2.4.1.1 Use of an Average Annual Growth Rate	31
2.2.4.1.2 Use of a percentage of change	33
2.2.4.2 Calculation of the observed evolution with all values	34
2.2.4.3 Methods comparison	35
2.3 Synthesis	37
2.3.1 Assessment method considerations	37
2.3.1.1 The methods are not mutually exclusive	37
2.3.1.2 Indicators not evolving in a clear manner	37
2.3.1.3 Use of averages	37
2.3.1.4 Historical path of the indicator	37
2.3.1.5 Use of scenarios	37
2.3.2 Linear/non-linear description of the observed and desired evolution	37
2.3.2.1 Category 1 methods (target value and target year available)	38
2.3.2.2 Category 4 methods (trend assessment)	39

2.3.3	Significance and relevance of the observed evolution	40
2.3.3.1	Category 1 methods (target value and target year available)	41
2.3.3.2	Category 4 methods (trend assessment)	41
2.3.3.2.1	Rate or amount of change as a threshold	41
2.3.3.2.2	Use of significance tests	42
2.3.4	Which method should be used?	42
3.	Communication of indicator-based assessment	43
3.1	Context	44
3.2	Communication forms and target audiences	44
3.3	Communication of the process and results of indicator-based assessment	45
3.3.1	Communication at the level of individual indicators	45
3.3.1.1	Communication of category 1 methods (target value and target year available)	45
3.3.1.2	Communication of category 2 methods (target value available, no target year defined)	47
3.3.1.3	Communication of category 4 methods (trend assessment)	48
3.3.1.4	Summary and discussion	50
3.3.2	Communication at the level of groups of indicators (dashboards)	52
3.3.3	Communication at an aggregated level (aggregated dashboards)	52
3.3.3.1	Simple, non-interactive aggregated dashboards	54
3.3.3.2	Interactive aggregated dashboards for sustainable development	54
3.4	On the use of symbols	59
4.	Conclusions and Outlook	61
4.1	Interactions	62
4.1.1	Frameworks and models of interactions are needed	62
4.1.2	Common classifications of data is a prerequisite	63
4.2	Improving the communication of indicator-based assessment	63
4.3	Role of indicator-based assessment in the post Rio+20 agenda	63
4.4	Role of indicator-based assessment outside the field of sustainable development	64
	Bibliography	65
	Annexes	67

Introduction

Context of this handbook

This handbook on indicator-based assessment is anchored in the work carried out on sustainable development indicators (SDIs) by several European countries and international organisations. Following the recommendations of the Agenda 21 adopted in 1992 in Rio de Janeiro (UN 1992) ⁽²⁾, these countries and organisations started elaborating systems or sets of SDIs to monitor sustainable development in the mid-90s. This was done in order to provide a reliable information basis for a large audience, from individuals to decision-makers.

Several of these actors were soon confronted with the difficulty of communicating the sometimes complex information provided by these indicator systems. This difficulty results from three characteristics of sustainable development indicators and indicator sets or systems:

- Firstly, the SDI sets or systems typically entail a large number of indicators given the multidimensional character of sustainable development. This characteristic conflicts with the communication function of indicators and with the fifth of the Bellagio principles (practical focus): ‘a limited number of indicators or indicator combinations to provide a clearer signal of progress’ ⁽³⁾.
- Secondly, for the user to understand and interpret the indicators, this same multidimensional character implies that he has to grasp numerous concepts that he is not familiar with a priori.
- Thirdly, the sustainable development indicators are generally classic economic, social or environmental indicators, which are presented in the perspective of sustainable development, outside their original context.

The seventh Bellagio principle ⁽⁴⁾ (effective communication) recommends a simple and fast reading of an indicator or group of indicators. This is not facilitated by the last two characteristics mentioned above. For these reasons, over the last decade, several countries as well as Eurostat were faced with the challenge of delivering simple messages that are easy to understand for the general public, policy-makers and/or stakeholders. They started almost simultaneously developing methods to synthesize the information given by indicators, and/or for assessing the progress towards sustainable development. The core of this synthesis activity called ‘indicator-based assessment’ consists in the following steps that are detailed in the Terminology section of Chapter 2:

- 1) Define a desired evolution for each indicator, on the basis of the frame of reference.
- 2) State the observed evolution of the indicator.
- 3) Compare the observed evolution with the desired evolution. Depending on the result of the comparison (the observed evolution is better, similar or worse than the desired evolution), the indicator is attributed to a positive, negative or neutral category (or any intermediate category between ‘positive’ and ‘negative’).

These parallel developments of indicator-based assessment methods were however not coordinated. They have consequently led to numerous approaches, which are not quite comparable a priori. They have also caused a certain confusion about the terms and concepts that are used.

Furthermore, the descriptions of the indicator-based assessment methods that are applied by national institutions (national statistical offices, ministries or public organisations) are often not published, at least not in detail, nor have they been presented to peers from national institutions. The methods could therefore not be reviewed by peers.

This situation raises questions among the users, sometimes even within the statistical offices that are applying the methods of indicator-based assessment. It also complicates the task for official statistics agencies or for ministries that wish to benefit from previous experiences in order to apply indicator-based assessment to their own indicators, either for the measurement of sustainable development or for another domain.

Finally, indicator-based assessment methods are used to improve the communication and the diffusion of SDIs and to draw the attention of policy-makers, the civil society and the media. However, the organisations who developed these methods are aware that their use raises several issues, such as the risk of delivering simplistic messages and the risk of misinterpretation of the results by the users. These organisations wonder whether indicator-based assessment is the way forward and have expressed their need for an analysis of the strengths and weaknesses of indicator-based assessment.

Given these conditions, an Expert Group on indicator-based assessment was set up in the context of the Eurostat Working Group on Sustainable Development Indicators (WG SDI). It consists of experts from Eurostat, the United Nations Economic Commission for Europe (UNECE) and eight European countries (Belgium, France, Germany, Netherlands, Norway, Sweden, Switzerland, United Kingdom) that have relevant expertise with indicator-based assessment methods.

Aim and structure of this handbook

The objective of the handbook is to present and discuss the current state of the art in Europe in the field of indicator-based assessment in official statistical institutions, ministries and public organisations. The handbook aims to share and compare experiences, create a common base of knowledge and unify the

⁽²⁾ Chapter 40.

⁽³⁾ These principles, and their updated version, the BellagioSTAMP (Bellagio Sustainability Assessment and Measurement Principles), serve as guidelines for assessing progress towards sustainable development. (Hardi, Zdan 1997 and IISD 2009).

⁽⁴⁾ Also mentioned as the sixth principle in the BellagioSTAMP.

terminology in this field of activity, but not to provide any recommendation on which assessment method should be used. It does not claim to answer the question whether indicator-based assessment is the way forward, however it contributes to the debate on this matter. This overview of the situation should give a basis for further, potentially common steps. This handbook is aimed at people from administrations or civil society working with lists or systems of indicators, not only SDIs, and willing to improve the communication on the indicators. It could also contribute to the work on Sustainable Development Goals (SDGs) initiated after the Rio+20 Conference in 2012 (UN 2012, paragraph 250 ff), especially in the definition of indicators and targets and in the use of methods to assess the indicators and communicate the results of the assessments.

This handbook is neither about the measurement of sustainable development ⁽⁶⁾, nor on the assessment of the relevance of indicators to measure sustainable development or some aspects of it, nor on the building of indicator systems in general.

This handbook describes the goals and the role of indicator-based assessment as well as the context in which it is inserted. It suggests a common terminology in the field of indicator-based assessment, establishes a classification of the methods and compares those used by the members of the Expert Group on the basis of a common sample of indicators. This handbook also gives an overview of how the methods and results of the indicator-based assessment are communicated. It finally outlines possible future developments in the field of indicator-based assessment.

NOTE ON THE TERMINOLOGY

Terminology has been defined in the context of this handbook. It is presented in Chapter 2, section 2.1.

The terms that are defined in this terminology are in colour at the beginning of each new section in Chapters 1 to 4.

⁽⁶⁾ For this purpose see e.g. the Conference of European Statisticians recommendations on measuring sustainable development (UN 2014a).

The role of indicators and assessment



1.1 Policy context

The work on sustainable development indicators (SDIs) was launched in 1992 at the United Nations Conference on Environment and Development (also known as the Rio Earth Summit). Governments recognised in Agenda 21 that indicators of sustainable development need to be developed to provide a solid basis for decision-making at all levels and to contribute to a self-regulating sustainability of integrated environment and development systems (Agenda 21, Chapter 40, paragraph 40.4). This document also encouraged the use of such indicators in satellite accounts and eventually in national accounts and mentioned that the development of indicators needs to be pursued by the UN Statistical Office, as it draws upon evolving experience in this regard (Agenda 21, Chapter 40, paragraph 40.6). Two decades later, considering the development of accounting systems (e.g. the System of Environmental-Economic Accounting SEEA (UN/EC/IMF/OECD/World Bank 2003)), several organisations recommend a slightly different approach, namely the usage of SDIs that are based on accounting systems because these indicators are able to measure assets and flows and thus have a strong connection to core elements of the Brundtland definition of sustainable development (see e.g. Hass, Palm 2012, p. 38 ff, or UN 2014a). The final document of the UN Conference on Sustainable Development Rio+20 held in 2012 reemphasizes the role and the importance of indicators to measure and accelerate progress (UN 2012, paragraph 104).

In order to support the implementation of this commitment, the UN Commission on Sustainable Development (UN CSD) adopted a work programme on sustainable development indicators at its third session in 1995. The objective of this work programme was to develop and define SDIs, to promote their use and enable countries to use SDIs. A set of more than 100 SDIs was defined and tested by more than twenty countries from different regions as well as Eurostat.

At the European level, Eurostat played a key role to promote the use of SDIs. After the United Nations World Summit on Sustainable Development (Johannesburg summit) in 2002, Eurostat first set up a task force on SDIs in 2002, gathering countries with experience in the development and use of SDIs. In 2005, this task force defined a first set of SDIs at the EU level. In 2006, a renewed sustainable development strategy was adopted at the EU level, and Eurostat received the mandate to publish a list of SDIs in a *monitoring* report every two years in cooperation with Member States. Hence, the task force was replaced by a working group open to all EU and EFTA countries in order to involve them in the drafting of the monitoring reports and to share experiences between Member States and with Eurostat.

When the work on SDIs started in the late 1990s, the challenge for the organisations working on SDIs was mainly to define a measurement framework, also called an indicator framework, and then select the relevant SDIs to monitor sustainable development as defined in the framework. There have been many attempts to create general frameworks for sustainable development indicators at regional, national and international

levels. A UN report (UNDESA 2007) gave a survey of indicator frameworks previously and currently applied, including the DPSIR⁽⁶⁾ frameworks, issue- or theme-based frameworks (typically policy-based approaches, see hereunder), capital frameworks, accounting frameworks and aggregated indicators.

Typically, the approaches that organisations followed to define their indicator framework can be classified into two categories: policy-based approaches and conceptual approaches (adapted from UN 2009).

a) Policy-based approaches

Several countries or international organisations use their sustainable development strategy (prepared by policy-makers) as a *frame of reference* (see Chapter 2, section on Terminology) and hence defined indicators to monitor or assess the progress towards the objectives of the strategy. This is for example the case for the United Kingdom SDIs (see e.g. DEFRA 2013) or for the EU SDIs (see e.g. Eurostat 2011). In such cases, the SDI sets reflect the content of the strategies and are limited to the issues/themes covered by these strategies. They are typically organised according to the issues/themes of the strategy.

b) Conceptual approaches

Other countries or international organisations defined a *conceptual framework* (see Chapter 2, section on Terminology) that allows selecting and organising the indicators. The framework includes a frame of reference that can be independent from political priorities⁽⁷⁾, or that can be derived from a sustainable development strategy (e.g. Belgium). A conceptual framework also includes a model of the interactions between the economic, environmental and social factors and/or between flow and stock variables (e.g. DPSIR framework, or the Swiss MONET typology). In both cases, the resulting *indicator system* presents a structure that includes both a thematic and a model-based organisation of the indicators.

It is worth pointing out that the difference between the two described approaches lies in the measurement concept. The sustainable development strategies on which the policy-based approaches rely are, or can be, themselves based on references such as the Brundtland report or the Agenda 21. These references are, however, the basis for the definition of the issue of sustainable development and are not concepts for the *measurement* of sustainable development. Accordingly, a 'conceptual framework' does not only define what to measure (the themes, the topics), but also how to measure it by using a model of sustainable development processes and/or their interactions (e.g. an indicator typology).

⁽⁶⁾ Driving Forces – Pressures – State – Impact – Responses.

⁽⁷⁾ Such as the Swiss Sustainable Development Principles that take into account the themes, principles and objectives of the international sustainable development documents or commitments (e.g. Brundtland report, Agenda 21).

1.2 The need to provide a summary or a comprehensive assessment

As mentioned in the Introduction, a key challenge for the organisations publishing sets of SDIs is to deliver easily communicable synthetic messages on the progress towards sustainable development and hence to ensure that **indicators** are actually used in the policy process or by citizens. This is all the more important because, even if large resources have been allocated to establish SDIs, their effectiveness in influencing actual policies and practices often remains limited (see Parris, Kates 2003, Bell, Morse 2003, Pinter, Hardi, Bartelemus 2005) and there is in general a growing pressure to use highly aggregated indices instead of lists of SDIs since they look more attractive in terms of communication ⁽⁸⁾ (see Hass, Palm 2012, p. 22–23).

The work on SDIs is characterised by the normative nature of the concept of sustainable development. Sustainable development principles and objectives have been adopted at the United Nations Conference on Environment and Development in 1992 (Rio Earth Summit), in particular in the Rio Declaration and in Agenda 21. They are the reference for most actors working on sustainable development and were actually developed at regional, national, supranational levels according to the political priorities and responsibilities. This normative nature of sustainable development implies that the persons in charge of its monitoring could rely on the internationally agreed principles and objectives to go beyond the simple statistical description of the state and evolution of a phenomenon, and to assess the progress towards sustainable development.

This raises the question of the objectives against which the indicators are assessed. In the sets of indicators founded on a policy-based approach, the objectives against which the indicators are assessed are determined in the sustainable development strategy, be it at the regional, national, or supranational level. In the sets of indicators founded on a concept-based approach, the objectives against which the indicators are assessed can either be defined within the **conceptual framework** underlining the indicators (e.g. Switzerland), or taken from the sustainable development strategy or other policy documents promoting sustainable development (e.g. Belgium).

In order to carry out such assessments and deliver synthetic messages, the organisations in charge of SDIs have been working at two levels: at the level of each single indicator in a first stage, and at the level of groups of indicators in a second stage.

a) Assessment at the level of individual indicators

At the level of individual indicators, the organisations have been working on conveying a simple message on the basis of the trends shown by the indicators. In order to provide such

⁽⁸⁾ Indices aggregating a large number of variables or indicators in a single number (e.g. environmental performance indicators) are in general more attractive than lists of SDIs but they raise several other problems in terms of transparency and methodology.

a message, they compared the **observed evolution** and/or **status** of an indicator with its **desired evolution**. This comparison allows providing a positive, negative, neutral (or any intermediate category between ‘positive’ and ‘negative’) assessment of the indicator.

The desired evolution of an indicator may take the form of a quantified **target** with or without a year to reach it (e.g. achieve 3 % of GDP ⁽⁹⁾ in research and development expenditures by 2020), or of a general direction to be followed (e.g. reduce atmospheric pollution). The assessment method will change according to the form of the desired evolution (see Chapter 2).

The definition of the desired evolution of the indicators depends largely on the indicator framework chosen by the organisations publishing sets of SDIs:

- If the framework is policy-based, the desired evolutions are set by the objectives defined in the sustainable development strategy. For example in Germany, 38 indicators have been defined to monitor the National Strategy for Sustainable Development and for most of these indicators quantitative targets were defined.
- If the framework is conceptual, the desired evolutions can either be set by the content of the framework, or by political objectives defined in sustainable development strategies. For example, Switzerland has defined in its indicator framework several principles or postulates that define the **desired directions** for the indicators. In Belgium, although the indicators are defined on the basis of a conceptual framework, they are assessed against political objectives defined in various political commitments.

It is also important to specify that several SDI sets also include accompanying variables, named ‘contextual indicators’ ⁽¹⁰⁾ (or also ‘descriptive’ or ‘situational’ indicators), which are not assessed because they do not monitor a specific objective or are not policy responsive (e.g. number of persons per household in the EU SDI set) and their assessment is not considered to be relevant.

b) Assessment at the level of groups of indicators

At the level of groups of indicators (in the form of lists or tables), the main task consists in summarizing the assessments made at the level of the individual indicators. This task raises the following questions: How to summarize/synthesize the individual assessments? And, How to communicate the summary/synthesis? The main concern when answering these questions is to maintain a high level of transparency, avoiding to hide some negative trends, for example, and in some way facilitating and encouraging the reader to look at

⁽⁹⁾ Gross Domestic Product.

⁽¹⁰⁾ Generally, contextual indicators are difficult to interpret in a normative way. They are included in the set because they provide valuable background information on issues having direct relevance for sustainable development policies and are helpful to understand a topic (Eurostat 2011, p. 38). Considering the definition of ‘indicator’ that is used in this handbook, the ‘contextual indicators’ should rather be named contextual variables, or contextual information.

the individual indicators. The summary assessment should indeed invite the reader to look at the individual indicators and hence increase the accessibility to information provided by indicators. Chapter 3 deals in detail with these issues, showing several examples in some Expert Group countries.

1.3 Contribution of indicator-based assessment to policy evaluation

Evaluation can be defined as the ‘*judgement of interventions according to their results, impacts and needs they aim to satisfy*’ (European Commission 2004). Evaluation is a discipline in itself that has a very large spectrum. Since the work on indicator-based assessment uses indicators to evaluate or assess whether sustainable development objectives are achieved, it looks at the outcomes of policies and is therefore clearly connected to evaluation and in particular policy evaluation (see Figure 1.1).

SDIs indeed provide information on the global impacts, i.e. the long-term effects/outcomes of an intervention. Therefore, when we assess SDIs against political objectives, we actually look at these outcomes. We assess the evolution or *status* of indicators that can be indirectly influenced by the sustainable development policies or strategy, but also by any other phenomenon, be it of a social, economic or environmental nature, like a global economic crisis or a volcanic eruption. Hence, the work done on indicator-based assessment is part of an evaluation process, but is of course only a small contribution to an evaluation process that should cover all the elements mentioned in Figure 1.1.

1.4 Role of official statistics

Information produced by official statistical institutes differs from the myriad of other information available in that these institutes have to follow principles to produce and disseminate

statistical information. These are the United Nations Fundamental Principles of Official Statistics (UN 2014b), the European Statistics Code of Practice (ESSC 2011), and the guidelines and principles about quality in statistics established by many countries.

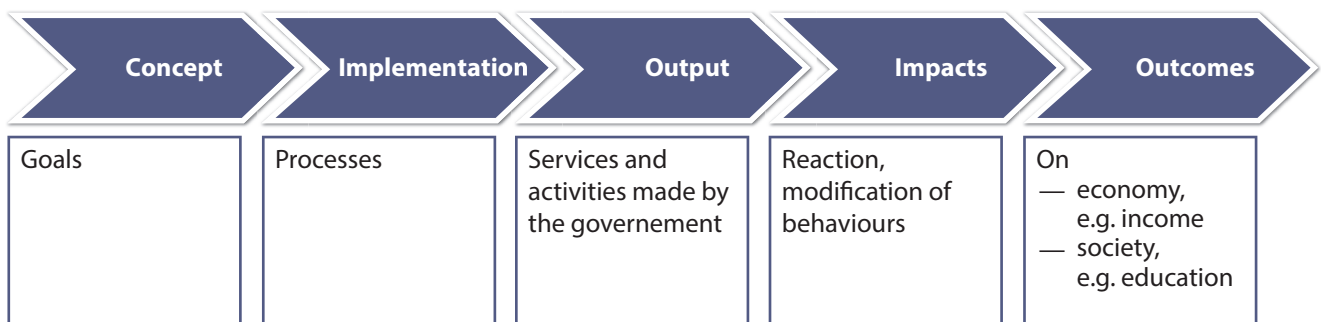
These principles play an important role in the very normative field of sustainable development and considering the risks mentioned in the Introduction (i.e. delivering simplistic messages and misinterpretation of the results by the users). They state among other things that:

- statistical authorities have to be professionally independent from other policy, regulatory or administrative bodies as well as from private sector operators in order to ensure the credibility of the statistics (Principle 1 of the EU principles);
- statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedure for the collection, processing, storage and presentation of statistical data (Principle 2 of the UN principles);
- statistics have to be produced and disseminated in an impartial and objective manner, respecting scientific independence (Principle 6 of the EU principles);
- statistics should be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance (Principle 15 of the EU principles).

Assessment methods and the assumptions made have to be defined absolutely independently from any political pressure, and have to be well documented and published.

Of course, the fundamental principles of official statistics can be adopted and applied by official and non-official bodies that are not official statistical agencies. But official statistical agencies are probably best suited to guarantee the independence from political pressure and the respect of strictly professional consideration because their independence is specified in law, according to Principle 1 of the EU principles.

Figure 1.1: General model of the policy evaluation. Adapted from Balthasar 2000 (p. 13–25) in Feller-Länzlinger et al. 2010.



**Methodology of
indicator-based assessment**

2

How to get messages across using **indicators**? This is the core question this handbook tries to answer by presenting state-of-the-art **indicator-based assessment** methods used in several European countries and Eurostat in the field of sustainable development. It is not the aim of this chapter to judge whether or not the methods are good: they all have their advantages and shortcomings. This chapter, rather, presents the rationale behind each of them and the situations in which they can be used, with examples from practice.

The chapter begins with the definition of the terminology used throughout the whole handbook and then presents the methods themselves, grouped into four categories. In order to do so, each method has been applied on the same indicators from the Eurostat Monitoring report (Eurostat 2011). Some of the methods are described in a very detailed manner in country-specific publications, some in less detail. In such cases, the information given in this chapter was communicated directly by the Expert Group members. A list of the country-specific publications and websites is available in Annex 2. The data aims at illustrating the methods with practical examples and is not the most up-to-date. Unless otherwise stated, EU-27 data downloaded from the Eurostat database between March and June 2012 was used.

2.1 Terminology

As mentioned in the Introduction, several countries and international organisations began developing and using sustainable development **indicators** (SDIs) and carrying out **indicator-based assessment** in the middle of the 1990s, following the recommendations of the Agenda 21. These activities were poorly coordinated, if at all, and this has led to a variety of assessment methods and terms used to describe these methods. The Expert Group on Indicator-based Assessment agreed on common terminology, presented below, and this is used throughout the whole handbook.

It should be noted that this terminology does not aim at providing universally valid definitions. The common understanding of terms such as ‘indicator’ or ‘frame of reference’ can be different outside the context of this handbook.

2.1.1 Terms terminology

If a term is in colour in the text of a definition or in the rest of the handbook, it means it is itself defined in the terminology.

Average Annual Growth Rate (AAGR)

The Average Annual Growth Rate ⁽¹⁾ is the constant growth rate needed to obtain the final value from the initial value in n periods. The formula for its calculation is shown in equation 1.

$$AAGR = \sqrt[n]{\frac{\text{final value}}{\text{initial value}}} - 1 \quad (1)$$

⁽¹⁾ The AAGR is sometimes also called Compound Annual Growth Rate (CAGR).

where n is the number of periods.

Base year

The base year is the first year of the **period analysed**. The value of the **indicator** at base year usually is the **reference value** (see Figure 2.1 and also the definition of ‘reference value’). It can be decided that, by principle, the base year is the same year for all indicators of a given indicator set or system, e.g. 1992 (year of the Earth Summit in Rio), or year of publication of a political programme such as a national sustainable development strategy. In practice, however, there is not always data available for each indicator at base year. In such cases, the earliest data after the base year is generally used.

It can also be decided that the base year is indicator-specific, e.g. the year of the setting of a **target** for a given indicator. However, in this case, adaptations sometimes have to be made in practice due to lacking data at base year.

Conceptual framework

A conceptual framework is essentially a measurement framework for the issue dealt with (e.g. sustainable development) that allows the setting up of an **indicator system**. It consists of the following elements:

- A **frame of reference**, which defines the important aspects for the highlighted issue (e.g. population in good health, economic stability or preservation of natural resources can be defined as important aspects for the issue of sustainable development). The frame of reference also allows the definition of the **desired evolution** of each **indicator**. Examples of a frame of reference are national sustainable development strategies, or the three dimensions and the 45 Principles of Sustainable Development in the Swiss SDI system (MONET).
- A systemic structure, which ensures that the important aspects are monitored by the relevant indicators and helps to avoid an arbitrary indicator selection. Examples of systemic structures are the DPSIR structure for the Belgian SDIs or the indicator typology in the Swiss indicator system.
- Predefined criteria and processes for indicator selection, so that the roles and responsibilities are set in advance for all actors involved with indicator selection.

Desired direction

A desired direction is set for any **indicator**, implicitly or explicitly (see definition of ‘Indicator’). The desired direction has three ‘modalities’: ‘increase’, ‘decrease’, or ‘stability’. If an indicator is accompanied by a **target value**, the desired direction consists of the direction the indicator should follow from the **reference value** in order to reach the target. If an indicator is not accompanied by a target, the desired direction is set, implicitly or explicitly, according to the **frame of reference**. The desired direction is an element of the desired evolution.

Desired evolution

The desired evolution is the evolution against which the **observed evolution** of the **indicator** is compared in order to perform **indicator-based assessment**. The desired evolution of an indicator can simply be its **desired direction** (increase, decrease, stability) or it can be calculated on the basis of additional information such as a **reference value**, a **target value**, and a **target year**.

Frame of reference

A frame of reference is intrinsically normative as it allows the definition of a **desired evolution** for each **indicator** of an **indicator set**. In practice, a frame of reference can be a political programme (such as a sustainable development strategy (SDS), or a part of a **conceptual framework** (such as the qualitative objectives and the principles of the MONET conceptual framework).

It should be noted that the notions of ‘frame of reference’ and ‘conceptual framework’ refer to the measurement by indicators of the issue dealt with (e.g. sustainable development). A political programme such as a sustainable development strategy, although it relies on conceptual aspects, cannot be considered to be a conceptual framework on its own, because it lacks a systemic structure as well as selection criteria and processes.

Indicator

The term ‘indicator’ is used in many different contexts and by various scientific disciplines. There is no universally valid definition of it. In the context of this handbook, an indicator is defined as a statistical variable presented in the form of a time series and chosen to monitor the evolution of a specific aspect of the issue dealt with (e.g. the aspects of health, economic efficiency or natural resources for the issue of sustainable development) towards a **desired direction** or a **target value**. This understanding of the term ‘indicator’ corresponds to what is also called ‘performance’ or ‘normative’ indicators. Usually, an indicator consists of only one variable (or is the result of the computation of several variables). The variable does not have to be causally linked to the process under review, but it should represent an important aspect of this process. The variable is always accompanied by information about what process it relates to, why it has been chosen to represent this process, in what direction it should evolve/what **target** it should reach, and in what direction it actually evolves/what its situation is regarding the target. It should be noted, however, that this information may not systematically be communicated to the user.

Indicator-based Assessment

Indicator-based assessment refers to the positive, negative, neutral (or any intermediate class between ‘positive’ and ‘negative’) qualification of an **indicator** based on the comparison between its **observed evolution** (and/or **status**), and the **desired evolution** set for the indicator by means of a **frame of reference**.

The name ‘indicator-based assessment’ originates from the desire of the Expert Group members to clearly distinguish this activity from any other assessment activity, such as the assessment of the statistical quality of indicators, or the assessment of the effectiveness of policies (controlling).

‘Indicator-based assessment’ consists in the following steps:

- 1) Define a desired evolution for each **indicator**, on the basis of the frame of reference.
- 2) Determine a method to compare the desired evolution with the observed evolution or status of the indicator, and define the **thresholds** to delimit what observed evolution is considered fully, partially and not corresponding to the desired evolution (resulting in a positive, neutral and negative qualification of the indicator).
- 3) Calculate the observed evolution or status of the indicator in the **period analysed**.
- 4) Compare the observed evolution of the indicator with the desired evolution using the method determined in step 2. Depending on the result of the comparison (observed evolution fully, partially or not corresponding to the desired evolution) and according to predefined thresholds, the indicator is attributed to a positive, negative or neutral class (or any intermediate class between ‘positive’ and ‘negative’).

In this process, it is important to distinguish between the desired (or observed) direction (increase, decrease or stability) and the result of the indicator-based assessment (positive, neutral or negative qualification of an indicator).

Indicator set

In this handbook, an indicator set is defined as a list of **indicators** based on a **frame of reference**, as opposed to an **indicator system**, in which the indicators rely on a **conceptual framework**.

Indicator system

In this handbook, an indicator system is understood as comprising a **conceptual framework** and the **indicators** stemming from such a conceptual framework (see also Chapter 1, section 1.1).

Monitoring

Monitoring includes the collection, analysis and presentation of information with the aim of tracking the evolution of a particular domain or field of activity of a public authority or society in a methodical manner, in the long run, for early detection of problematic changes (Wachter 2010).

Observed evolution

In order to carry out **indicator-based assessment**, the values of an **indicator** over time need to be summarized in an observed evolution. A method has to be defined to do so, which implies

choices such as the **period analysed** (e.g. from the year 2000 to the last available value of the indicator) and the values of the indicator taken into account. The observed evolution is therefore a simplification of the information. It can be expressed e.g. only as a direction ('increase', 'stability', 'decrease') in trend assessment methods (see section 2.2.4), or as a direction together with a rate of change (e.g. **AAGR**). The values of the indicator that are taken into account for the definition of the observed evolution influence the observed direction, and it is important that users of indicator-based assessment methods are aware of this.

Period analysed

The period analysed is the period considered for carrying out **indicator-based assessment** and has to be defined for each **indicator**. It begins with the **base year** (see Figure 2.1) and often ends, in practice, with the last available value of each indicator. A minimum length of the period and a minimum amount of data points in the period have to be defined, below which an indicator-based assessment is not possible. In practice, a minimum length of three years and a minimum amount of three data points are often required.

It is also possible to define two (or more) periods analysed, e.g. in order to carry out a short- and a long-term assessment. In the United Kingdom for example, two assessments are carried out: one over the period '1990 to last available value' and one over the period '2003 to last available value'. In addition, the direction of change between the penultimate and the latest year is provided as information (see DEFRA 2010).

Reference value

Value of the **indicator** at **base year** (see Figure 2.1). Some countries use an average of several years — usually the average of the value of the indicator at base year and of the two adjacent or following values — as a reference value at base year. They do so in order to reduce the effect of the value of a particular year.

Status

The status of an **indicator** is a value of the indicator that does not take into account the evolution of the indicator. For example: last available value of the indicator used for comparison with **target value** when no **target year** is defined.

Target Path

Path calculated on the basis of a **reference value** and of the **target value** at **target year** (see Figure 2.1). Two methods for calculating a target path are currently in use:

Linear target path

Path with a constant slope (straight line). The slope of the path is calculated with the following formula (equation 2):

$$\text{Slope of the linear target path} = \frac{\text{final value} - \text{initial value}}{\text{final year} - \text{initial year}} \quad (2)$$

Non-linear, exponential target path

Path with a constant percentage of increase (curve). The growth rate of the path is its **AAGR**.

Target value

A target, or target value, is understood in this handbook as a value that the **indicator** should reach, accompanied or not by a deadline to achieve this value (**target year**) (see Figure 2.1). Among the Expert Group members using targets (Belgium, Eurostat, Germany and Switzerland), the setting of targets is not the responsibility of the statistical office, but of policy-makers and/or politicians. A typical example of a target set by policy-makers with a target year is the Kyoto target for greenhouse gas emissions; a typical example of a target set by policy-makers with no target year is the share of official development assistance (ODA) in the gross national income (GNI).

Target year

Year at which a **target value** set for an **indicator** has to be reached (see Figure 2.1). The target year allows to calculate a **target path** or to extrapolate the indicator based on the value(s) of the indicator at some point in time and on the value of the target at target year. Not all target values are accompanied by a target year.

Threshold

Thresholds are needed to delimit the positive, negative and neutral qualification classes attributed to **indicators** by the **indicator-based assessment** process (comparison of the **observed** and **desired evolutions**). The definition of the thresholds depends on the type of assessment method chosen, in particular on how the desired and observed evolution are compared.

2.1.2 Charts terminology

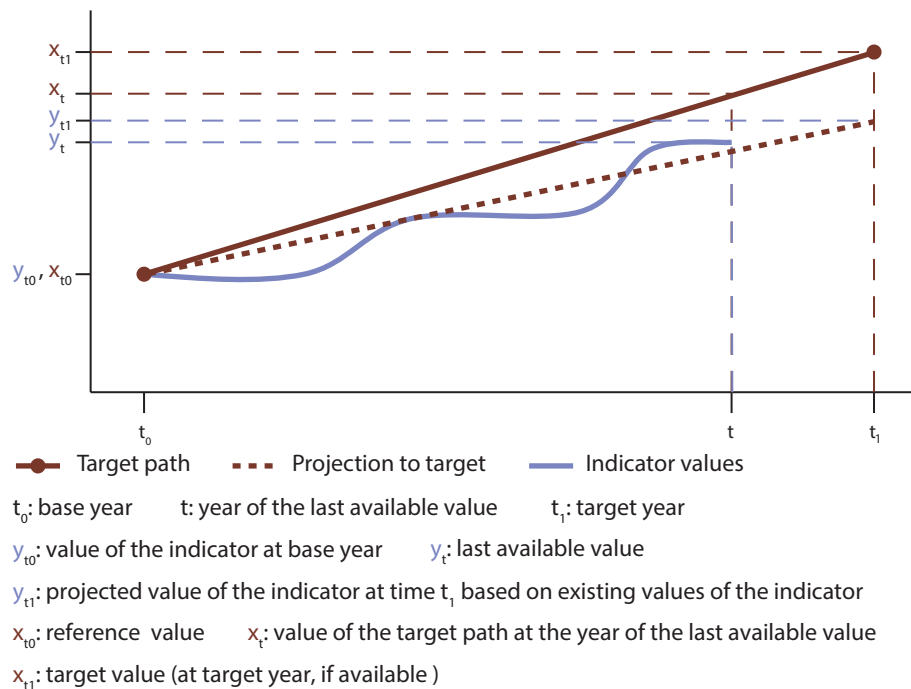
Figure 2.1 presents all the elements used in the different methods presented below. Time (t) is represented on the x-axis. The values of the indicator (y) and, where applicable, of the **target** and **target path** (x) are represented on the y-axis.

Not all these elements are relevant for all methods.

2.2 Method classification

The members of the Expert Group use a variety of methods to carry out **indicator-based assessment**, and there are further methods that could potentially be implemented. This chapter classifies the methods into four categories, on the basis of the (non-)availability of certain pieces of information concerning the **desired evolution** of the **indicators** (**target value**, year to achieve the target value, **desired direction**). This is crucial because in the current context, it is usually not the role of those responsible for selecting indicators and for carrying out indicator-based assessment to define target values and **target years**.

Figure 2.1: General chart terminology



They therefore depend on the setting of such targets by policy-makers. The types of methods that can be applied, greatly differ depending on the availability of such pieces of information. The suggested classification is presented in Figure 2.2.

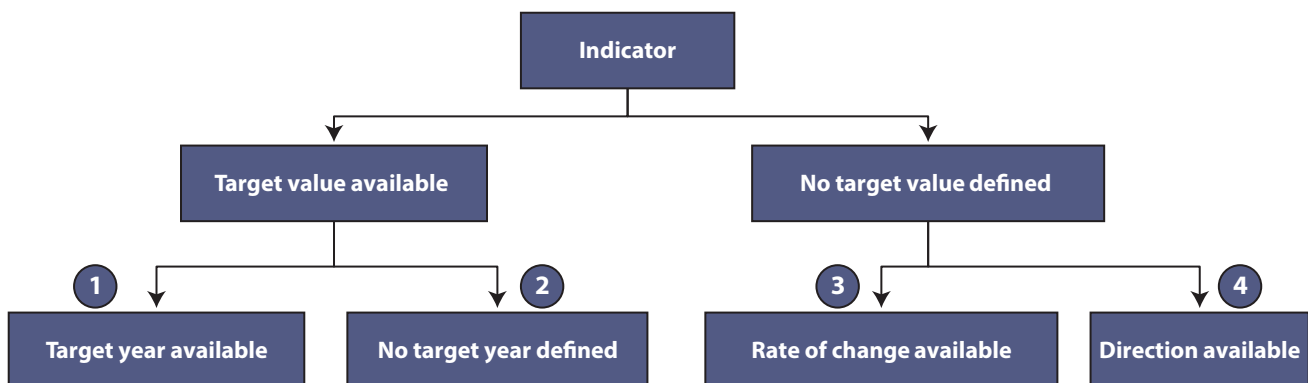
The numbering of the methods (1 to 4) refers to a hierarchy in the methods. The category 1 methods require the most pieces of information concerning the desired evolution of each indicator (target value and target year). The category 4 methods only require that a desired direction (increase, decrease or stability) is defined for each indicator. This means that if an indicator is accompanied by all the information required for the use of a category 1 method (target value and target year), it can also be assessed with a category 2, 3 or 4 method, and so on. It should be noted that these methods are not mutually exclusive.

The classification of the methods is inspired and adapted from the work carried out by Hulliger, Lussmann 2010. A correspondence table in Annex 1 shows the similarities between the classification used in this chapter and the one suggested by Hulliger & Lussman. Eurostat uses a further method to assess decoupling indicators, which is not presented in this handbook.

The sections 2.2.1 to 2.2.4 below describe the categories of assessment methods 1 to 4 (see numbers in Figure 2.2). They are structured in the following manner:

- General description of the method with an example of an EU sustainable development indicator that can be assessed with this kind of method, if relevant.

Figure 2.2: Classification of the methods into four categories



- List of the additional elements that need to be defined in order to be able to carry out indicator-based assessment. In practice, it is the responsibility of those carrying out indicator-based assessment to define these additional elements.
- For the assessment categories 1 and 4, sub-classification of the methods within the category, with comments on the similarities and differences between the methods, where relevant.
- Detailed description of the method(s) actually used by the members of the Expert Group.
- Comparison of the methods in use for the assessment categories 1, 2 and 4.
- Calculation method for the **target path** (see section 2.2.1.1) or the projection/extrapolation (see section 2.2.1.2).
- For example: slope or **AAGR** calculated on the basis of the reference value and of the target value.
- Actual value(s) of the indicator to be taken into account for the comparison with the target path or the target value.
- For example: reference value x_{t_0} and last available value of the indicator y_t .
- Method for comparing the value(s) of the indicator with the target path or the projected value at target year with the target value.

2.2.1 Category 1: target value and target year available

The methods described in this section rely on the existence of a relevant target, x_{t_1} defined at year t_1 (target year). A typical example of an indicator that could be assessed with this kind of method is the EU sustainable development indicator ‘Early leavers from education and training’: the EU target is to reach a maximum of 10 % of early leavers from education and training (x_{t_1}) by 2020 (t_1) in comparison to 2003 (t_0).

In addition to the target value x_{t_1} and the target year t_1 , the following elements have to be defined in order to carry out an indicator-based assessment:

- Base year t_0 (defined as 2003 in the above example).
- Reference value x_{t_0} (in practice, the reference value is often the value of the indicator at base year for category 1 assessment methods — except for the German method, see below).

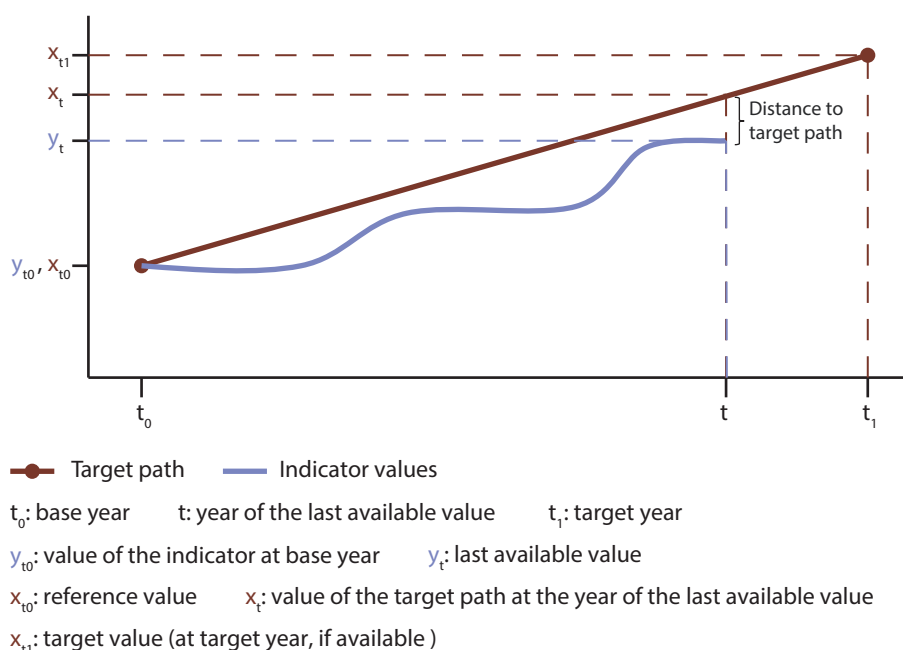
- For example: ratio of the slope of the target path and of the actual evolution of the indicator.
- **Threshold(s)** of deviation of the value(s) of the indicator from the target path or target value. These thresholds determine the assessment categories.
- For example: less than 20 % deviation from target path or from target value for a positive assessment.

In practice, the base year is often defined by policy-makers, but the other elements are usually defined by those responsible for the indicators and their assessment.

2.2.1.1 Target path

The assessment consists in comparing the indicator (e.g. difference between reference value and last available value) to a theoretical, calculated target path (see Figure 2.3). The target path is connecting the value of the indicator at base year (reference value) with the target value at target year. There are several

Figure 2.3: Comparison of a value of the indicator with a target path



ways to calculate a target path (e.g. linear and non-linear target paths), and these differences have an impact on the result of the assessment. Actually, the target path itself is rarely calculated and presented, it is rather its slope or growth rate that is used for the assessment.

When calculating a target path (or its slope/growth rate), the direction in which the indicator should evolve in order to reach the target (desired direction) is implicitly taken into account. The direction depends on the relation between the reference value and the target value: if the target value is higher than the reference value, the implicit desired direction will be 'increase'; if the target value is lower than the reference value, the implicit direction will be 'decrease'. The desired direction is indicated by the sign of the slope or of the growth rate.

The desired rate of change is also implicitly taken into account and corresponds to the slope or the growth rate of the target path.

2.2.1.1.1 Linear target path

A linear path from the base year to the target year is calculated with a simple linear equation:

$$x = x_{t_0} + (t_x - t_0) * \left(\frac{x_{t_1} - x_{t_0}}{t_1 - t_0} \right) \quad (3)$$

Where t_x is the year at which one wants to calculate x .

The linear path corresponds to the desired evolution.

In the methods in use presented below, the observed evolution is calculated on the basis of two values of the indicator: reference value (x_{t_0}), which corresponds to the value of the indicator at base year (y_{t_0}), and last available value of the indicator (y_t). These methods therefore do not consider the historical path of the indicator between t_0 and t .

Methods in use

Belgium ⁽²⁾

A linear target path starting at base year is calculated with the following formula for the theoretical slope:

$$\text{theoretical slope} = \frac{x_{t_1} - x_{t_0}}{t_1 - t_0} \quad (4)$$

The ratio of the distance between the reference value x_{t_0} and the last available value y_t versus the distance between the reference value x_{t_0} and the target path value x_t (obtained with equation 4) is then calculated and used for the assessment:

$$\text{ratio used by Belgium} = \frac{y_t - x_{t_0}}{x_t - x_{t_0}} \quad (5)$$

There are two thresholds: ratio of 0.8, and of 0.5. There are hence three assessment classes:

- If the ratio is greater than 0.8, which means that the indicator is evolving in the desired direction and that the last available value of the indicator amounts to 80 % or more of the target path value, the assessment is positive.
- If the ratio is smaller than 0.8 but greater than 0.5, which means that the indicator is evolving in the desired direction and that the last available value of the indicator lies between 50 % and 80 % of the target path value, the assessment is neutral.
- If the ratio is smaller than 0.5, which means that the indicator either evolves in the desired direction but the last available value of the indicator is smaller than 50 % of the target path value, or the indicator evolves in the direction opposite to the desired direction, the assessment is negative.

Switzerland

The ratio of the observed and desired slopes is calculated directly:

$$\text{ratio used by Switzerland} = \frac{\frac{y_t - x_{t_0}}{t - t_0}}{\frac{x_{t_1} - x_{t_0}}{t_1 - t_0}} \quad (6)$$

There are two thresholds: ratio of 0.97, and ratio of 0. There are hence three assessment classes:

- If the ratio is greater than 0.97, which means that the indicator is evolving in the desired direction and that the last available value of the indicator amounts to 97 % or more of the target path value, the assessment is positive.
- If the ratio is smaller than 0.97 but greater than 0, which means that the indicator is evolving in the desired direction and that the last available value of the indicator lies between 0 and 97 % of the target path value, the assessment is neutral.
- If the ratio is smaller than 0, which means that the indicator is evolving in the direction opposite to the desired direction, the assessment is negative.

Similarities and differences between the Swiss and Belgian methods

Belgium and Switzerland actually calculate the same ratio: Belgium obtains the value of the target path at the year of the last available value (x_t) by applying the theoretical slope to the reference value at base year (x_{t_0}). This is explained in the following equation:

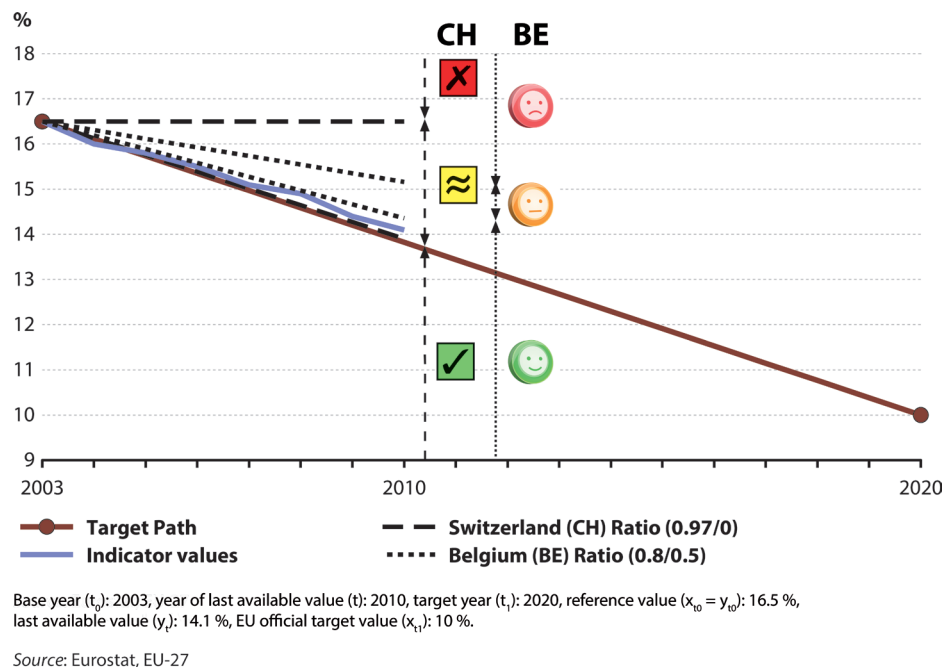
$$x_t = x_{t_0} + (t - t_0) * \left(\frac{x_{t_1} - x_{t_0}}{t_1 - t_0} \right) \quad (7)$$

When replacing x_t by its equivalent in the equation 5 (ratio used by Belgium), the result is the ratio used by Switzerland.

It is interesting to note that the thresholds used are not the same (0.8 and 0.5 for Belgium, 0.97 and 0 for Switzerland). This of course has an influence on the assessment results: the

⁽²⁾ Belgium changed its assessment method in 2011 for a method that is similar to the Eurostat method (for indicators with a target value and target year). However, in order to show as many different examples as possible in this handbook, it was decided to describe the method used until 2011.

Figure 2.4: Differences in the assessment results for linear target path methods using different thresholds (Belgian and Swiss methods) for the indicator ‘Early leavers from education and training’ (see also Table 2.2a)



threshold used by Switzerland to get a positive assessment is more severe than the one used by Belgium (0.97 vs. 0.8). On the other hand, the threshold for a negative assessment is also lower for Switzerland (0) than for Belgium (0.5). An example is shown with the indicator ‘Early leavers from education and training’ (see Figure 2.4 and Table 2.2a).

2.2.1.1.2 Non-linear target path

A non-linear path from the base year to the target year is calculated. The non-linear target path currently in use is calculated with an AAGR (see equation 1).

Paths to target calculated on the basis of an AAGR have an exponential form represented by the following equation:

$$x = a * e^{b*t_x} \quad (8)$$

Where t_x is the year at which one wants to calculate x , and a and b are constants.

The target path corresponds to the desired evolution.

In the method in use presented below, the observed evolution is calculated on the basis of two values of the indicator: reference value (x_{t_0}) and last available value of the indicator (y_t). This method therefore does not consider the historical path of the indicator between t_0 and t .

Methods in use

Eurostat

The observed AAGR on the basis of the reference value at base year and the last available value is calculated as well as the theoretical AAGR on the basis of the reference value and the target value:

$$\text{Observed AAGR} = \sqrt[t-t_0]{\frac{y_t}{x_{t_0}}} - 1 \quad (9)$$

$$\text{Theoretical AAGR} = \sqrt[t_1-t_0]{\frac{x_{t_1}}{x_{t_0}}} - 1 \quad (10)$$

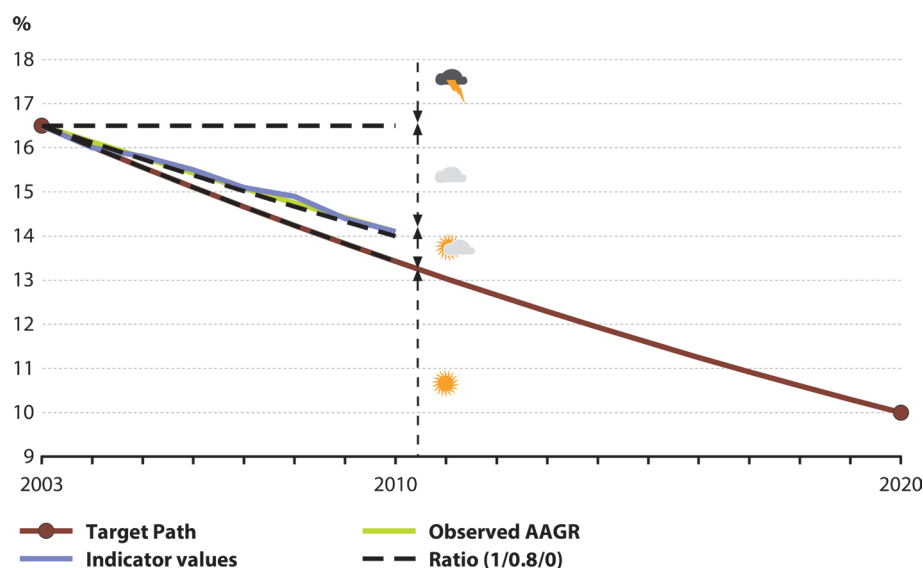
$$\text{Ratio} = \frac{\text{Observed AAGR}}{\text{Theoretical AAGR}} \quad (11)$$

There are three thresholds: ratio of 1, ratio of 0.8, ratio of 0, and hence four assessment classes⁽¹³⁾:

- If the ratio is greater than 1, which means that the indicator is evolving in the desired direction and that the observed rate of change of the indicator amounts to 100 % or more of the desired rate of change, the assessment is positive.

⁽¹³⁾ Eurostat changed its thresholds in 2013 and they are now identical to the thresholds used by Germany (0.95, 0.8 and 0 instead of 1, 0.8 and 0). Eurostat also now takes into account the results of the Spearman's rank correlation, as does Germany. The assessment is expressed using weather symbols which are solid if the trend is significant and shaded if this is not the case. However, in order to show as many different examples as possible in this handbook, it was decided to describe the method used until 2013.

Figure 2.5: Assessment results for a non-linear target path method (Eurostat method) for the indicator ‘Early leavers from education and training’ (see also Table 2.2a)



Base year (t_0): 2003, year of last available value (t): 2010, target year (t_1): 2020, reference value ($x_{t_0} = y_{t_0}$): 16.5 %, last available value (y_t): 14.1 %, EU official target value (x_{t_1}): 10 %.

Source: Eurostat, EU-27

- If the ratio is smaller than 1 but greater than 0.8, which means that the indicator is evolving in the desired direction and that the observed rate of change of the indicator lies between 80 % and 100 % of the desired rate of change, the assessment is moderately positive.
- If the ratio is smaller than 0.8 but greater than 0, which means that the indicator is evolving in the desired direction and that the observed rate of change of the indicator lies between 0 % and 80 % of the desired rate of change, the assessment is moderately negative.
- If the ratio is lower than 0, which means that the indicator is evolving in the direction opposite to the desired direction, the assessment is negative.

A concrete example of the use of this method is shown for the indicator ‘Early leavers from education and training’ (see Figure 2.5 and Table 2.2a).

2.2.1.2 Statistical projection: Extrapolation

An indicator time series can be projected to target year and assessed against the target value. The projected value at target year y_{t_1} is compared with the target value x_{t_1} (see Figure 2.6).

As is the case when calculating a target path, the desired direction and rate of change are implicitly taken into account, the former corresponding to the sign of the slope or of the growth rate of the projection, the latter to the slope or growth rate itself.

In the method presented below, the desired evolution cannot be summarized easily. It corresponds to the direction that the indicator should follow in order to reach the target (increase, decrease, or possible stability), and to a given rate or annual amount of change so that the target is reached at target year.

Methods in use

Germany

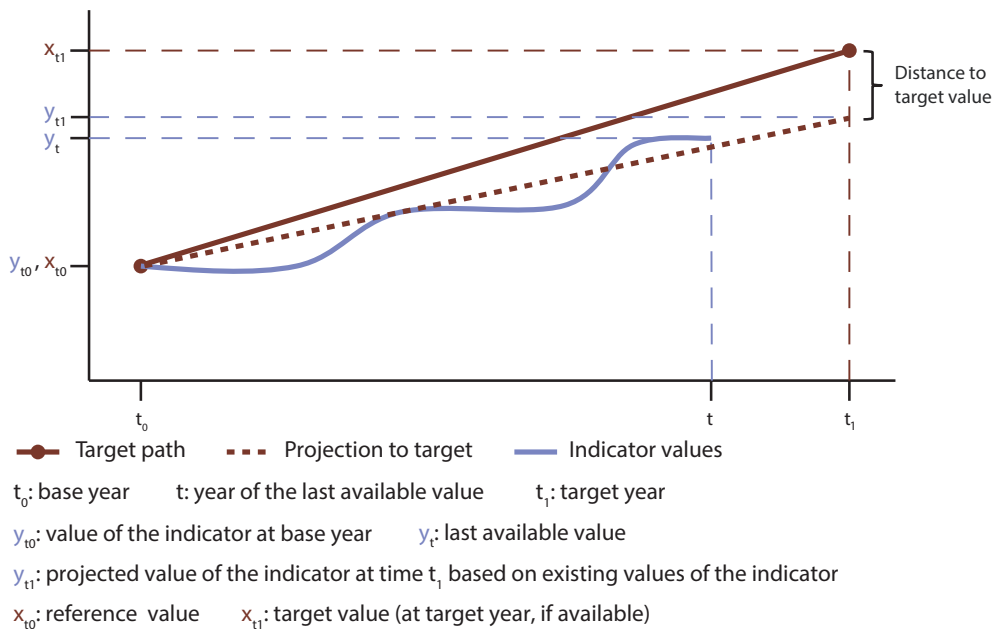
The AAGR of the last five years is calculated on the basis of the last available value y_t and of the value of the indicator five years before the last available value (y_{t-5}). Therefore, the projection actually does not begin at base year (which is a fixed year, set according to the target value), but at a year that changes each time there is a new value for the indicator (see equation 12).

$$\text{Observed AAGR over the five last available values of the indicator} = \sqrt[t-(t-5)]{\frac{y_t}{y_{t-5}}} - 1 \quad (12)$$

The indicator is then projected until target year t_1 on the basis of the calculated AAGR, the projection beginning at the value of the indicator five years before the last available value (y_{t-5}) (see equation 13).

$$y_{t_1} = y_{t-5} * \left[1 + \left(\sqrt[t-(t-5)]{\frac{y_t}{y_{t-5}}} - 1 \right) \right]^{t_1-(t-5)} \quad (13)$$

The ratio of the distance between the reference value x_{t_0} and the projected value y_{t_1} versus the distance between the reference value x_{t_0} and the target value x_{t_1} is then calculated and used for the assessment:

Figure 2.6: Comparison of a projected value with the target value

$$\text{Ratio} = \frac{y_{t_1} - x_{t_0}}{x_{t_1} - x_{t_0}} \quad (14)$$

In addition, a Spearman's rank correlation coefficient is calculated for the last five available values of the indicator (those used for the calculation of the AAGR). The formula for the Spearman's rank correlation coefficient r_s is the following:

$$r_s = \frac{\sum_i (c_i - \bar{c})(d_i - \bar{d})}{\sqrt{\sum_i (c_i - \bar{c})^2 \sum_i (d_i - \bar{d})^2}} \quad (15)$$

Where c_i is the rank of the years and d_i is the rank of the indicator values.

There are three **thresholds**: ratio of 0.95, ratio of 0.8, ratio of 0, and hence four assessment classes:

- If the ratio is greater than 0.95, which means that the indicator is evolving in the desired direction over the last five years and that the projected value at target years amounts to 95 % of the target value, the assessment is positive.
- If the ratio is lower than 0.95 but greater than 0.8, which means that the indicator is evolving in the desired direction over the last five years and that the projected value at target year lies between 80 % and 95 % of the target value, the assessment is moderately positive.
- If the ratio is lower than 0.8 but greater than 0, which means that the indicator is evolving in the desired direction over the last five years and that the projected value at target year lies between 0 % and 80 % of the target value, the assessment is moderately negative.

- If the ratio is lower than 0, which means that the indicator evolves in the direction opposite to the desired direction, the assessment is negative.

The result of the Spearman's rank correlation coefficient is used to tell whether the trend of the indicator over the five last available years is significant or not.

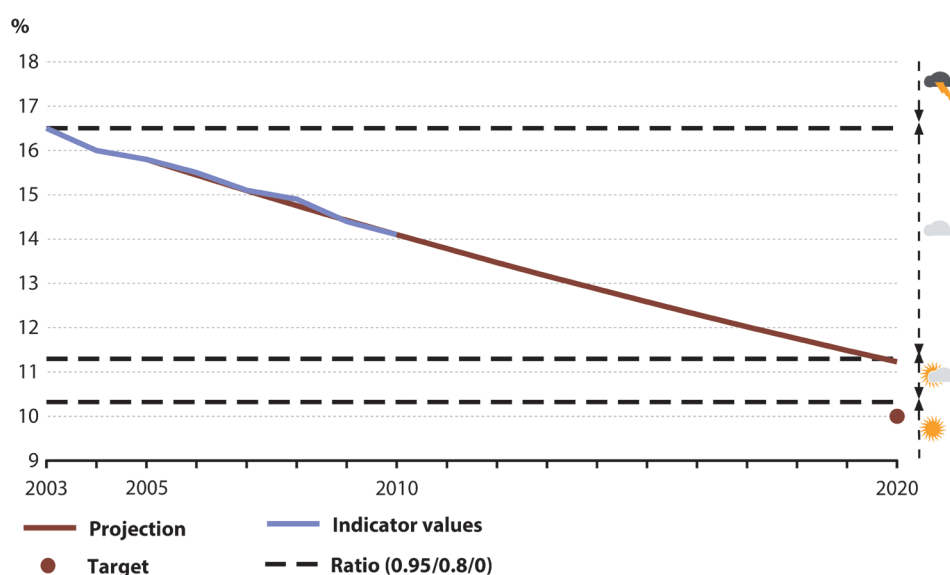
A concrete example of the use of this method is shown for the indicator 'Early leavers from education and training' (see Figure 2.7 and Table 2.2a).

2.2.1.3 Methods comparison

The main characteristics that differentiate the category 1 methods from one another are presented in Table 2.1.

The methods presented above have been tested on a common set of **indicators** from the Eurostat Monitoring report (Eurostat 2011) (see Introduction of this Chapter). As can be expected, not all methods deliver the same results, even if they are applied on the same indicators. The dispersion of the assessment results for each indicator can be determined by calculating the variance of the results. To calculate this variance, each class of assessment is expressed by a value without dimension (–1 for a negative assessment, –0.5 for a moderately negative assessment, 0 for a neutral assessment, 0.5 for a moderately positive assessment and 1 for a positive assessment). The variance is then calculated on these values. The variance of each indicator is represented in Figure 2.8. If the variance is high for a particular indicator, it means that the results of the different assessment methods applied on this indicator are quite different from one another. If the variance is low, it means that the assessment methods deliver similar results for a given indicator.

Figure 2.7: Assessment results for a projection to target value method (German method) for the indicator ‘Early leavers from education and training’ (see also Table 2.2a)



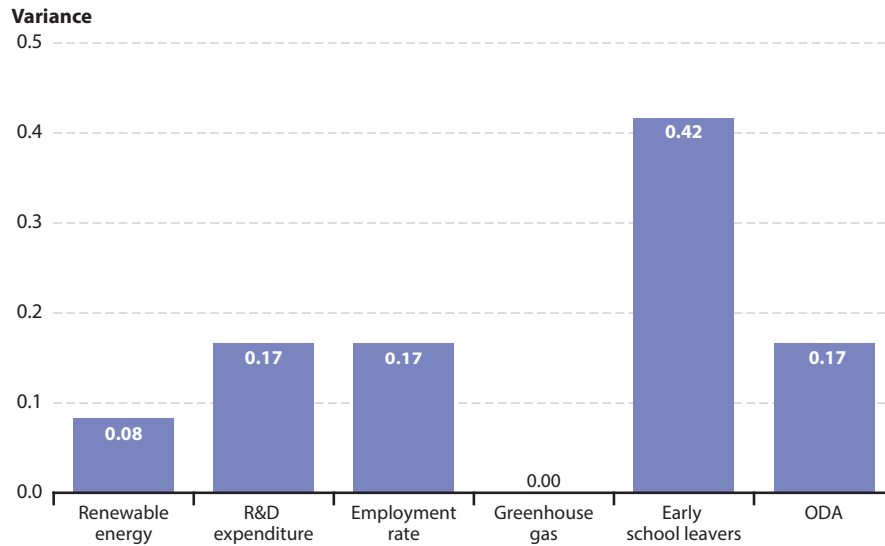
Base year (t_0): 2003, year of last available value (t): 2010, target year (t_1): 2020, reference value ($x_{t_0} = y_{t_0}$): 16.5 %, y_{t_1} : 15.8 %, last available value (y_t): 14.1 %, EU official target value (x_{t_1}): 10 %.

Source: Eurostat, EU-27

Table 2.1: Comparison of the category 1 methods in use

	Values of the indicator taken into account		Shape of the target path or extrapolation (in brackets: calculation method)	Assessment classes	Thresholds
	First value (reference value)	Last value			
Belgium	Same as Eurostat	Same as Eurostat	Linear (slope)	Three: - 'positive' - 'neutral' - 'negative'	>0.8 (positive) >0.5 and <0.8 (neutral) <0.5 (negative)
Eurostat ⁽¹⁴⁾	Reference value for the target (e.g. value of the indicator in 2003 for the indicator 'Early leavers from education and training' if the EU target is used)	Last available value of the indicator	Exponential (AAGR)	Four: - 'positive' - 'moderately positive' - 'moderately negative' - 'negative'	>1 (positive) >0.8 and <1 (moderately positive) >0 and <0.8 (moderately negative) <0 (negative)
Germany	Reference value for the target (same as Eurostat) Projection: Value of the indicator five years before the last available value	Last available value of the indicator	Exponential (AAGR)	Four: - 'positive' - 'moderately positive' - 'moderately negative' - 'negative' + Spearman's rank correlation coefficient	>0.95 (positive) >0.8 and <0.95 (moderately positive) >0 and <0.8 (moderately negative) <0 (negative)
Switzerland	Same as Eurostat	Same as Eurostat	Linear (slope)	Three: - 'positive' - 'neutral' - 'negative'	>0.97 (positive) >0 and <0.97 (neutral) <0 (negative)

⁽¹⁴⁾ Eurostat changed its thresholds in 2013 and they are now identical to the thresholds used by Germany (0.95, 0.8 and 0 instead of 1, 0.8 and 0). Eurostat also now takes into account the results of the Spearman's rank correlation, as does Germany. The assessment is expressed using weather symbols which are solid if the trend is significant and shaded if this is not the case. However, in order to show as many different examples as possible in this handbook, it was decided to describe the method used until 2013.

Figure 2.8: Dispersion of the results of the category 1 assessment methods in use, tested on six indicators

Full names of the indicators as in the Eurostat Monitoring report (Eurostat 2011):

Renewable energy: "Electricity generated from renewable sources"

R&D expenditure: "Total R&D expenditure"

Employment rate: "Total employment rate"

Greenhouse gas: "Greenhouse gas emissions"

Early school leavers: "Early leavers from education and training"

ODA: "Official development assistance as share of gross national income"

Source: Eurostat, EU-27

The strongest variance can be observed for the indicator 'Early leavers from education and training'. Figure 2.4 in section 2.2.1.1.1 shows the difference that is caused by different thresholds only (comparison between the Belgian and Swiss method that are identical except for the thresholds). Further sources of variance are due not only to differences between the

thresholds, but of course also to differences between the methods. The different results obtained by the different methods on the indicator 'Early leavers from education and training' are summarized in Table 2.2a. The variance for the indicator 'Greenhouse gas emissions' is zero because all the methods tested deliver the same result (i.e. 'positive', see Table 2.2b).

Table 2.2: Differences in the results of the category 1 methods in use**a) for the indicator 'Early leavers from education and training'**

	Base year	Target year	Target value	Ratio	Assessment
Belgian method	2003	2020	10 %	0.90	(see Figure 2.4)
Eurostat method				0.76	(see Figure 2.5)
German method				0.81	(see Figure 2.7)
Swiss method				0.90	(see Figure 2.4)

b) for the indicator 'Greenhouse gas emissions'

	Base year	Target year	Target value	Ratio	Assessment
Belgian method	2000	2020	80 %	1.62	
Eurostat method				1.58	
German method				2.40	
Swiss method				1.62	

NB: the ratios are not calculated on the basis of the same elements (see the descriptions of the methods in use in sections 2.2.1.1 and 2.2.1.2).

2.2.2 Category 2: target value available, no target year defined

The methods described in this section rely on the existence of a relevant **target**, x (see Figure 2.9). A typical example of an **indicator** that could be assessed with this kind of method is the indicator ‘Official development assistance as share of gross national income’, which is included in most SDI sets or systems. The UN target is to reach a minimum of 0.7 % of the gross national income used for official development assistance, with no year defined to reach this share ⁽¹⁵⁾. Of course, any indicator for which a **target value** and a **target year** are available can be assessed according to this method, by simply not taking into account the fact that a target year is available.

In the case that an indicator is equipped with a target value without a target year (x), an assessment is possible provided that the following elements are defined:

- Value(s) of the indicator taken into account for comparing with the target value.
- For example: last available value of the indicator y_t .
- Method for comparing the value(s) of the indicator with the target value.
- For example: comparison of the value(s) of the indicator and of the target value.
- **Thresholds** of deviation of the value(s) of the indicator from the target value. These thresholds determine the assessment classes.

- For example: target value as a single threshold, the value(s) of the indicator either below or above the target value.

Methods in use

Germany

Germany developed two methods for indicators with a target value but without a year defined to achieve the target value:

1. Simple comparison of the last available value (y_t) with the target value (x) taking into account the **desired direction**.

There is one threshold, the target value (x), and hence two assessment classes:

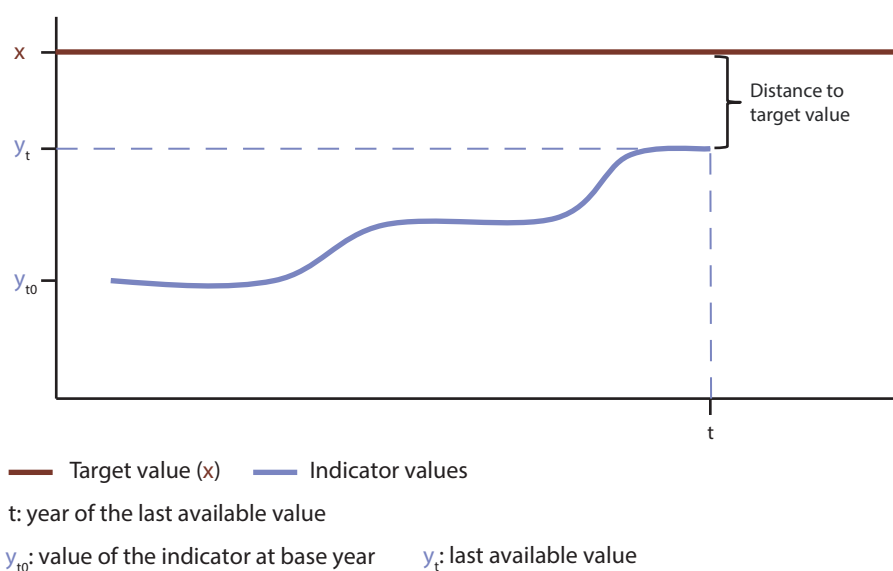
- a) If the desired direction is ‘increase’ and y_t is higher than x , or if the desired direction is ‘decrease’ and y_t is lower than x , the assessment is positive.
- b) If the desired direction is ‘decrease’ and y_t is higher than x , or if the desired direction is ‘increase’ and y_t is lower than x , the assessment is negative.

A concrete example of the use of this method is shown for the indicator ‘Official development assistance as share of gross national income’ (see Figure 2.10 and Table 2.3).

In this example, the implicit desired direction is ‘increase’. With the assessment method presented above, the indicator would be assessed positively if its last available value would be higher than 0.7 %, and negatively otherwise.

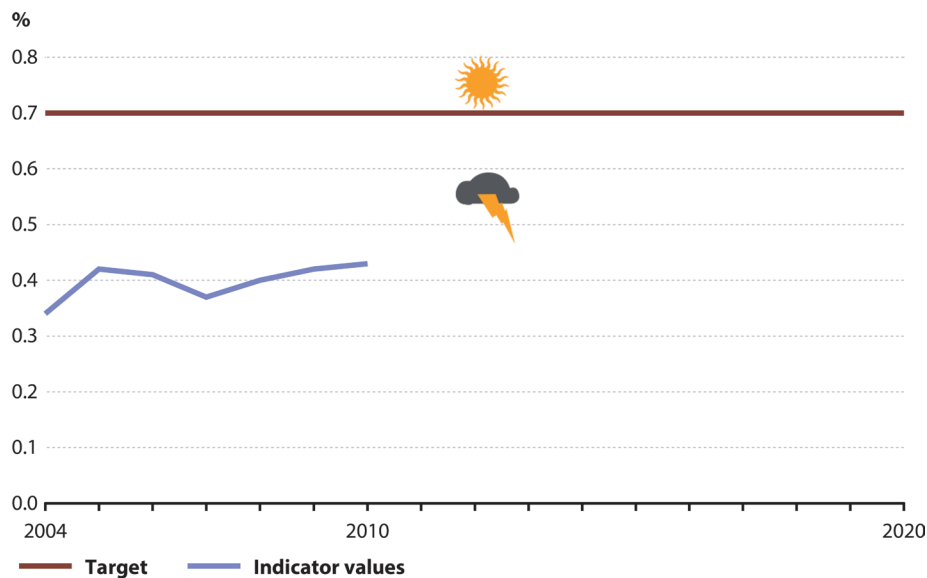
This method is also referred to as a ‘status assessment’ (as opposed to ‘trend assessment’, section 2.2.4).

Figure 2.9: Comparison of a value of the indicator with the target value



⁽¹⁵⁾ Intermediate targets and target years have been defined at the EU level. However, the UN target setting was chosen for the purpose of using a well-known example.

Figure 2.10: Assessment results for a category 2 method, simple comparison with target value (German method) for the indicator ‘Share of official development assistance in gross national income’ (see also Table 2.3)



Source: Eurostat, EU-27

Last available value (y_t): 0.43 %, EU official target value (x): 0.7 %.

Source: Eurostat, EU-27

2. Year of reaching of the target

The AAGR of the last five years is calculated on the basis of the last available value y_t and of the value of the indicator five years before the last available value (y_{t-5}). The year at which the target would be reached provided that the indicator continues to evolve with this AAGR is calculated with the following formula:

$$\text{year of reaching of the target} = \frac{\ln\left(\frac{x}{y_t}\right)}{\ln(1+AAGR)} + t \quad (16)$$

The calculated year is compared with the year of the last available value plus ten years and the result of the comparison is attributed to one of the three following classes:

- If the indicator evolves in the desired direction and the calculated year is smaller than the year of the last available value plus ten years, the assessment is positive.
- If the indicator evolves in the desired direction and the calculated year is larger than the year of the last available value plus ten years, the assessment is moderately negative.

- If the indicator evolves in the direction opposite to the desired direction, the assessment is fully negative.

Actually, this method implies the calculation of a (moving) target year. This target year is however not given by policy-makers, but is part of the assessment method.

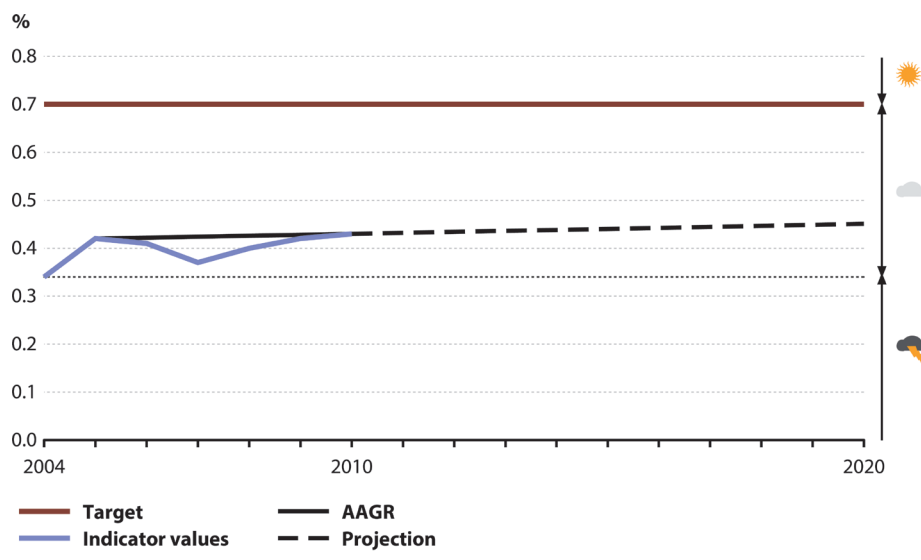
A concrete example of the use of this method is shown for the indicator ‘Official development assistance as share of gross national income’ (see Figure 2.11 and Table 2.3).

It should be noted that the AAGR and projection curves in Figure 2.11 look linear because the AAGR is very low, but that these curves are exponential.

2.2.2.1 Methods comparison

Table 2.3 shows the results obtained by the two methods presented above.

Figure 2.11: Assessment results for a category 2 method, year of reaching a target value (German method) for the indicator ‘Share of official development assistance in gross national income’ (see also Table 2.3)



Source: Eurostat, EU-27
 Last available value (y_t): 0.43 %, value of the indicator five years before the last available value (y_{t-5}): 0.42 %, EU official target value (x): 0.7 %.
 Source: Eurostat, EU-27

Table 2.3: Comparison of the category 2 methods in use for the indicator ‘Share of official development assistance in gross national income’

	Desired direction	Target value	Last value	AAGR (five last years)	Achieving target year	Assessment
Germany 1	↗	0.7 %	0.43	–	–	☁️ (see Figure 2.10)
Germany 2			0.43	0.47 %	2114	☁️ (see Figure 2.11)

2.2.3 Category 3: no target value defined, rate of change available

There may be cases where the **desired evolution** is an increase by a certain amount per year, but with no specific **target value** to reach. In other words, what is defined for the **indicator** is that it should evolve (indefinitely) with a certain slope or a rate of change (see Figure 2.12). No typical example of an indicator that can be assessed according to this method category can be given. Indeed, to our knowledge, this kind of method has never been implemented, at least in the field of the European or national SDIs. It is, however, mentioned here in order to show a possible alternative to the assessment methods in use.

An example of a required slope would be: decrease by 10 units each year (amount of change is constant, proportion of the successive values of the indicator changing).

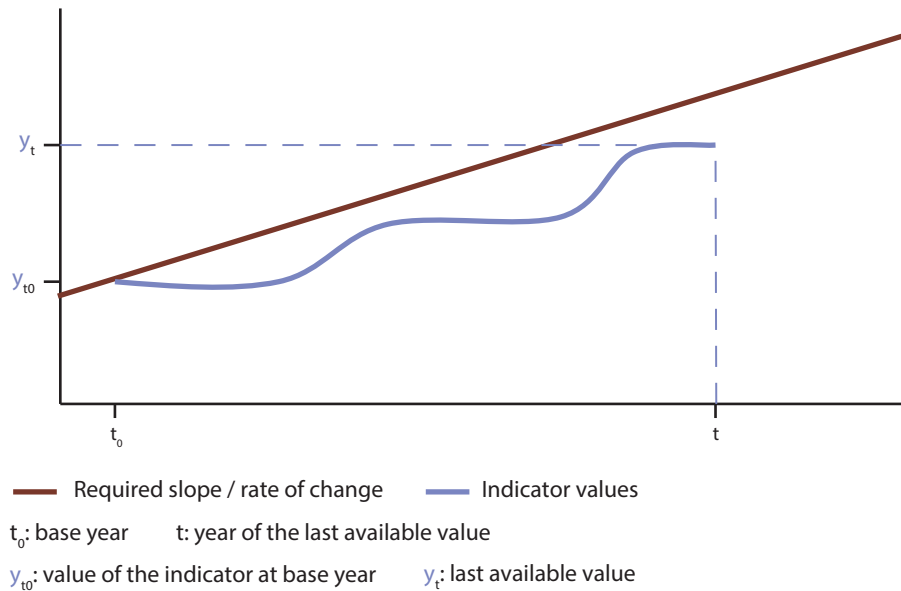
An example of a rate of change would be: decrease by 10 % each year (proportion of change is constant over time, amount of change varies).

An assessment method based on a required slope/rate of change implies that the following elements are defined:

- Values of the indicator to be taken into account to calculate the indicator’s slope/rate of change.
- For example: y_{t_0} and y_t .
- Method for comparing the slope/rate of change of the indicator with the required slope/rate of change.
- For example: ratio between the observed and required slope or rate of change.
- **Thresholds** of deviation of the slope/rate of change of the indicator from the required slope/rate of change. These thresholds determine the assessment classes.

Methods in use

No member of the Expert Group is currently using an assessment method based on a required rate of change.

Figure 2.12: Comparison of the slope or rate of change of the indicator with the desired slope or rate of change

2.2.4 Category 4: no target value defined, direction available (trend assessment)

In the case an indicator is only accompanied by a desired direction (no target value or rate of change available), it is possible to compare the observed direction of the indicator (derived from its **observed evolution**) with the desired direction (see Figure 2.13). Among the EU SDIs, most indicators are not accompanied by a target value, and no indicator is accompanied by a growth rate. There are therefore a lot of indicators that are currently assessed with a trend assessment method. Moreover, any indicator for which a target value and a **target year** or a growth rate are available can of course also be assessed according to a category 4 method, by simply not taking into account the fact that these elements are available.

In addition to the desired direction, the following elements need to be defined in order to carry out this type of assessment method:

- Period taken into account for the assessment (**period analysed**).

Examples:

- year of first available value to year of last available value of each indicator

- fixed starting year for all indicators to be assessed (e.g. milestone such as 1992 — UN earth summit in Rio) to year of last available value of the indicator

- fixed starting and ending years for all indicators to be assessed (e.g. 1990–2010).

- Values of the indicator taken into account in the defined period for calculating the observed evolution.

Examples:

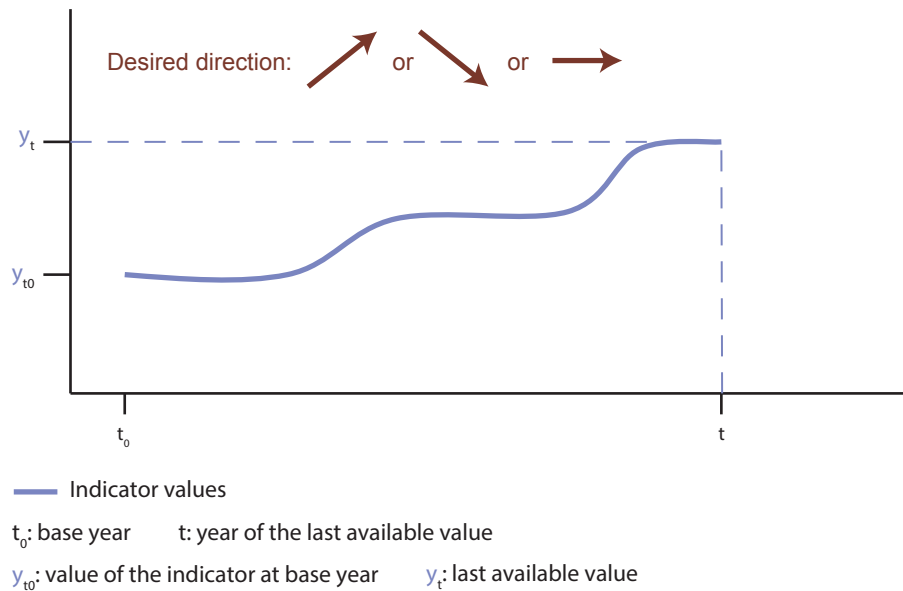
- first and last available values of the indicator in the period analysed
- average of the first value of the period and of the two neighboring values (y_{t_0-1} , y_{t_0} , y_{t_0+1}), and average of the last three available values (y_{t-2} , y_{t-1} , y_t) of the indicator
- all values of the indicator.

- Method for calculating the observed evolution.

Examples:

- percentage of change
- AAGR
- linear regression.

- **Thresholds** of deviation of the evolution of the indicator from the desired direction.

Figure 2.13: Comparison of the direction of the indicator with the desired direction

2.2.4.1 Calculation of the observed evolution with initial and final values

This approach uses the final value of the period (y_t) and the initial value (y_{t_0}) (expressed either as a single value or an average of values over a number of years) to determine the direction of change of the indicator (observed direction).

As such, this method does not consider the historical path of the indicator between t_0 and t .

2.2.4.1.1 Use of an Average Annual Growth Rate

The formula for the calculation of the AAGR is the following (equation 17):

$$AAGR = \sqrt[t-t_0]{\frac{y_t}{y_{t_0}}} - 1 \quad (17)$$

Methods in use

Belgium

The AAGR is calculated with the first and the last available values as y_{t_0} and y_t , respectively.

There are two thresholds: 1 % in the desired direction, 1 % opposite to the desired direction, and hence three assessment classes:

- If the sign of the AAGR (observed direction) corresponds to the desired direction and the AAGR is greater than 1 % in absolute terms, the assessment is positive.
- If the sign of the AAGR (observed direction) is opposite to the desired direction and the AAGR is greater than 1 % in absolute terms, the assessment is negative.

- If the AAGR is smaller than 1 % in absolute terms, the change is not considered to be significant and the assessment is neutral.

Eurostat ⁽¹⁶⁾

The AAGR is calculated the same way as for Belgium.

There are three thresholds: more than 1 % in the desired direction, more than 1 % opposite to the desired direction, between 0 and 1 % in the desired direction, between 0 and 1 % opposite to desired direction, and hence four assessment categories:

- If the sign of the AAGR (observed direction) corresponds to the desired direction and the AAGR is greater than 1 % in absolute terms, the assessment is positive.
- If the sign of the AAGR (observed direction) is opposite to the desired direction and the AAGR is greater than 1 % in absolute terms, the assessment is negative.
- If the sign of the AAGR (observed direction) corresponds to the desired direction and the AAGR is between 0 and 1 % in absolute terms, the assessment is moderately positive.
- If the sign of the AAGR (observed direction) is opposite to the desired direction and the AAGR is between 0 and 1 % in absolute terms, the assessment is moderately negative.

⁽¹⁶⁾ Starting from 2013, Eurostat also takes into account the results of the Spearman's rank correlation. The assessment is expressed using weather symbol which are solid if the trend is significant and shaded if this is not the case. However, in order to show as many different examples as possible in this handbook, it was decided to describe the method used until 2013.

Germany

The AAGR is calculated the same way as for Belgium.

There is only one threshold for the assessment (AAGR = 0), and hence only two assessment categories:

- If the AAGR is different from 0 and its sign (observed direction) corresponds to the desired direction, the assessment is positive.
- If the sign of the AAGR is opposite to the desired direction, the assessment is negative.

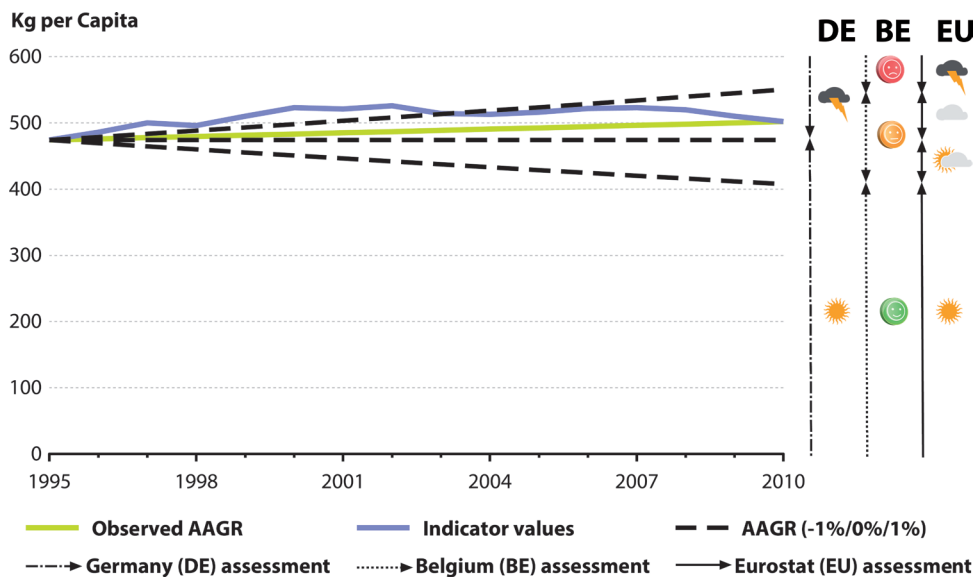
As additional information, the Spearman's rank correlation coefficient is calculated (see Equation 15 in section 2.2.1.2) and its value is used to tell whether the trend is significant or not.

The value of the coefficient is however not included in the result of the assessment itself, it is communicated separately.

Similarities and differences between the Belgian, Eurostat and German methods

The role of the thresholds and of the number of assessment classes is important as illustrated by the EU sustainable development indicator 'Municipal waste generation' (see Figure 2.14 and Table 2.4). For the German method, the value is higher than the 0 % threshold and the assessment is negative. Furthermore, the trend is significant in regard to the correlation test. For the Belgian and Eurostat methods, the AAGR is situated between 0 % and 1 % and the assessment is respectively neutral and moderately negative.

Figure 2.14: Differences in the assessment results for trend assessment methods using an AAGR for the calculation of the observed evolution for the indicator 'Municipal waste generation' (see also Table 2.4)



Base year (t_0): 1995, year of last available value (t): 2010, reference value ($x_{t_0} = y_{t_0}$): 474 Kg/capita, last available value (y_t): 502 Kg/capita.

Source: Eurostat, EU-27

Table 2.4: Differences in the results of the trend assessment methods using an AAGR for the calculation of the observed evolution for the indicator 'Municipal waste generation'

	Desired direction	Initial value	Final Value	AAGR	Assessment
Belgian method	↓	1995: 474	2010: 502	0.4 %	☹️ (see Figure 2.14)
Eurostat method					☁️ (see Figure 2.14)
German method					☹️⚡️ (see Figure 2.14)

2.2.4.1.2 Use of a percentage of change

The formula for the calculation of a percentage of change is the following:

$$\text{percentage of change} = \frac{y_t - x_{t_0}}{x_{t_0}} * 100 \tag{18}$$

Methods in use

Switzerland

The reference value (x_{t_0}) corresponds to the average of the value of the indicator at base year (y_{t_0}) and of the two neighboring values (y_{t_0-1} , y_{t_0+1}). The last available value is replaced by the average of the last three available values (y_{t-2} , y_{t-1} , y_t) (y_{tcalc}). The percentage of change is calculated with y_{tcalc} and x_{t_0} .

There are two thresholds: 3 % in the desired direction, 3 % opposite to the desired direction, and hence three assessment classes:

- a) If the sign of the percentage of change (observed direction) corresponds to the desired direction and the percentage of change is greater than 3 % in absolute terms, the assessment is positive.
- b) If the sign of the percentage of change (observed direction) is opposite to the desired direction and the

percentage of change is greater than 3 % in absolute terms, the assessment is negative.

- c) If the percentage of change is smaller than 3 % in absolute terms, the change is considered as not significant and the assessment is neutral.

United Kingdom

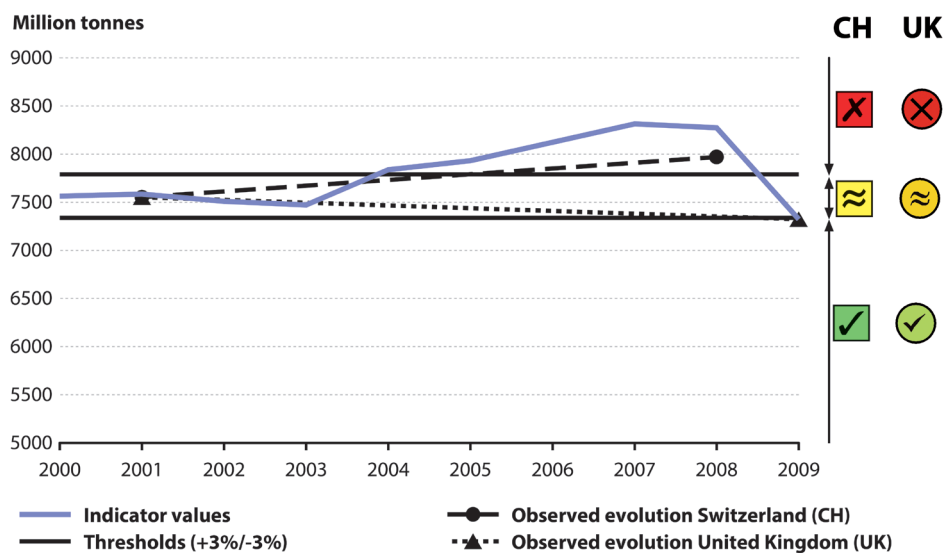
The reference value (x_{t_0}) corresponds to the average of the value of the indicator at base year (y_{t_0}) and of the two neighboring values (y_{t_0-1} , y_{t_0+1}). The percentage of change is calculated with the last available value of the indicator (y_t) and the reference value (x_{t_0}).

The thresholds and assessment classes are the same as for Switzerland ⁽¹⁷⁾.

Similarities and differences between the UK and the Swiss methods

The two methods are identical except for one thing: the Swiss method calculates the observed evolution by using the average of the last three available values, whereas the UK method uses the last available value only. This can have a strong influence on the result of the assessment, as shown with the EU SDI 'Domestic material consumption' (see Figure 2.15 and Table 2.5). The decrease observed in 2009 is important enough to have

Figure 2.15: Differences in the assessment results for trend assessment methods using a percentage of change for the calculation of the observed evolution for the indicator 'Domestic material consumption'





Base year (t_0): 2000, year of last available value (t): 2009, reference value (x_{t_0} = average (y_{t_0-1} , y_{t_0} , y_{t_0+1})): 7 553 348 ktonnes, last available value (UK: y_t = 7 324 683 ktonnes, CH: y_{tcalc} = average (y_{t-2} , y_{t-1} , y_t)): 7 971 592 ktonnes).

Source: Eurostat, EU-27

⁽¹⁷⁾ Note about the setting up of thresholds in the United Kingdom: It is interesting to note that prior to the 3 % threshold being introduced, indicator assessments were made by individual statisticians, without any consistent rules being followed. There was a call for a more consistent approach (not least by a Parliamentary Committee) even though the assessments were assumed to be individually statistical robust. Various methods and thresholds were tested, with the view to finding a consistent method that would produce results that were as close as possible to the assessments previously made individually. It was found that by the application of the 3 % threshold, the results most closely emulated the previous determinations across all the indicators.

Table 2.5: Differences in the assessment results for trend assessment methods using a percentage of change for the calculation of the observed evolution for the indicator 'Domestic material consumption'

	Desired direction	Initial value	Final Value	Percentage of change	Assessment
Swiss method	↘	Average 1999–2001: 7 553 348	Average 2007–2009: 7 971 592	5.5 %	 (see Figure 2.15)
UK method		Value 2009: 7 324 683		– 3.0 %	 (see Figure 2.15)

a positive assessment with the UK method. Due to the calculation of a final average, the influence of the 2009 value is lessened in the Swiss method.

2.2.4.2 Calculation of the observed evolution with all values

This approach uses all values within the period taken into account to determine the direction of change of the indicator (observed direction). It therefore considers the historical path of the indicator between t_0 and t . This kind of approach makes use of a model, with a number of underlying assumptions.

Methods in use

The Netherlands

An Ordinary Least Squares regression is calculated for the period analysed (between t_0 and t), where time is used as the explanatory variable and the values of indicator are used as the response variable. The slope is tested to know if it is significantly different from zero or not (95 % confidence interval).

The MS Excel Analysis Tool Pack can be used (Data Analysis > Linear Regression) ⁽¹⁸⁾. In the output, the slope corresponds to the Coefficient of 'X variable 1' (as it is in Excel) and the 95 % confidence interval is delimited by the values 'Lower 95 %' and 'Upper 95 %' for the 'X variable 1' (as it is in Excel).

Table 2.6: Comparison of the category 4 methods in use

	Values of the indicator taken into account		Method for calculating the observed evolution	Assessment classes	Thresholds
	First value (reference value)	Last value			
Belgium	Same as Eurostat	Same as Eurostat	AAGR	Three: - 'positive' - 'neutral' - 'negative'	1 %
Eurostat ⁽¹⁹⁾	First available value of the indicator in the period analysed	Last available value of the indicator in the period analysed	AAGR	Four: - 'positive' - 'moderately positive' - 'moderately negative' - 'negative'	0 % 1 %
Germany	Same as Eurostat	Same as Eurostat	AAGR	Two: - 'positive' - 'negative' + Spearman's rank correlation coefficient	0 %
Netherlands	All values of the indicator in the period analysed		Ordinary Least Squares regression with testing of the slope	Three: - 'positive' - 'neutral' - 'negative'	Significance of the slope of the regression
Switzerland	Average of the first available value of the indicator and of the two neighbouring values	Average of the last three available values of the indicator	Percentage of change	Three: - 'positive' - 'neutral' - 'negative'	3 %
United Kingdom	Same as Switzerland	Same as Eurostat	Percentage of change	Three: - 'positive' - 'neutral' - 'negative'	3 %

⁽¹⁸⁾ As in Microsoft Office Excel 2007.

⁽¹⁹⁾ Starting from 2013, Eurostat also takes into account the results of the Spearman's rank correlation. The assessment is expressed using weather symbol which are solid if the trend is significant and shaded if this is not the case. However, in order to show as many different examples as possible in this handbook, it was decided to describe the method used until 2013.

There are two thresholds: slope significantly different from zero and in the desired direction, slope significantly different from zero and opposite to the desired direction, and hence three assessment classes:

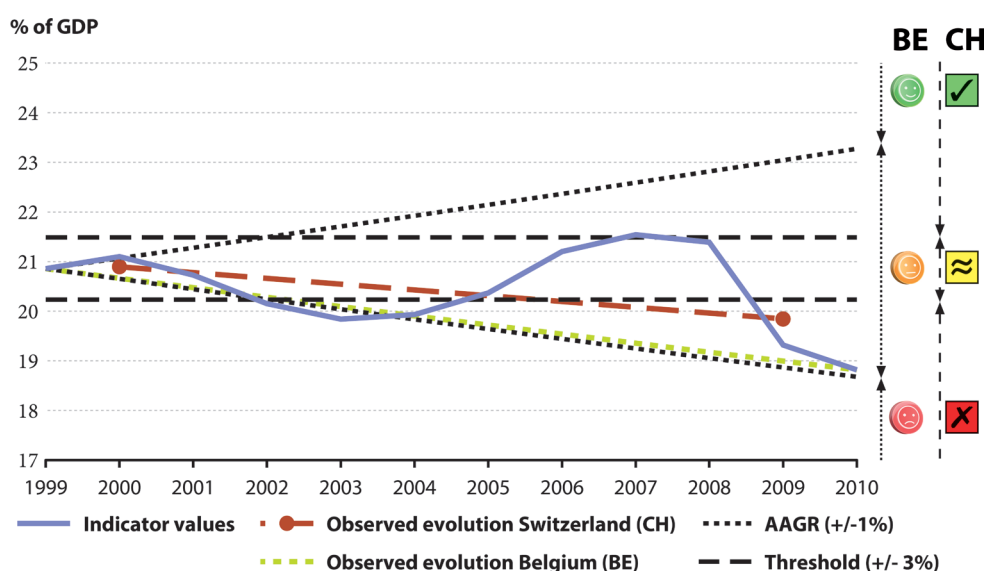
- a) If the slope is significantly different from zero and the observed direction (sign of the slope) corresponds to the desired direction, the assessment is positive.
- b) If the slope does not significantly differ from zero, the assessment is neutral (whatever the sign of the slope).
- c) If the slope is significantly different from zero and the observed direction (sign of the slope) is opposite to the desired direction, the assessment is negative.

2.2.4.3 Methods comparison

The main characteristics that differentiate the methods from one another are presented in Table 2.6.

The Belgian and the Swiss methods are used to compare methods calculating AAGR and methods calculating a percentage of change as they have the same number of assessment classes. The example of the EU SDI 'Total investment' (see Figure 2.16 and Table 2.7) shows that, for the same dataset, the Belgian method assesses the trend as neutral and the Swiss as negative. It also demonstrates that for a relatively long period of analysis (12 years), the fixed threshold of -3% is easier to exceed than the 'evolving' threshold of -1% AAGR. Using a -1% AAGR calculated over a 12 years period is similar to setting a fixed -10.5% threshold.

Figure 2.16: Differences in the assessment results for trend assessment methods using an AAGR or a percentage of change for the calculation of the observed evolution (Belgian and Swiss methods) for the indicator 'Total investment' (see also Table 2.7)



Base year (t_0): 1999, year of last available value (t): 2010,
 reference value (BE: $x_0 = 20.9\%$, CH $x_0 = \text{average}(y_{10-1}, y_{10}, y_{10+1}) = 20.9\%$),
 last available value (BE: $y_t = 18.8\%$, CH: $y_{\text{calc}} = \text{average}(y_{t-2}, y_{t-1}, y_t) = 19.8\%$).

Source: Eurostat, EU-27

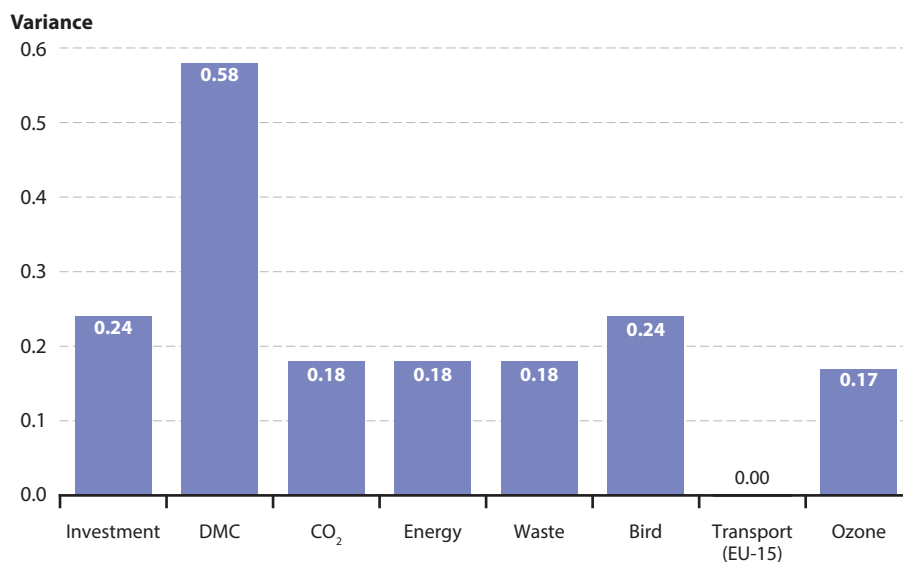
Table 2.7: Differences in the assessment results for trend assessment methods using an AAGR or a percentage of change for the calculation of the observed evolution (Belgian and Swiss methods) for the indicator 'Total investment'

	Desired direction	Initial value	Final Value	AAGR / % of change	Assessment
Belgian method	↗	Value 1999: 20.9	Value 2010: 18.82	-0.9%	☺ (see Figure 2.16)
Swiss method		Average 1999–2001: 20.9	Average 2008–2010: 19.8	-5.0%	☹ (see Figure 2.16)

It is difficult to compare the Dutch method with the others as it uses all the values of the time series by calculating a linear regression. It would however be interesting to do some further research to test the sensitivity of the Dutch method to extreme values in the time series and to compare the results with the other category 4 methods. One could also test different values for the confidence interval determining the significance of the slope for the Dutch method. Indeed, the values set for the confidence interval actually act as thresholds between the three assessment classes 'negative', 'neutral' and 'positive'.

As for category 1 methods, the trend assessment methods presented above have been tested on a common set of indicators from the Eurostat monitoring report (Eurostat 2011) over the period 1995–2010 (or shorter depending on data availability), and the dispersion of the assessment results has been calculated for each indicator (see section 2.2.1.3 for more details on the calculation of the dispersion). The dispersion is presented for each indicator in Figure 2.17. Table 2.8 shows the detail of the assessment results used for the calculation of the dispersion for each indicator and each method.

Figure 2.17: Dispersion of the results of the category 4 assessment methods tested on eight indicators



Full names of the indicators as in the Eurostat Monitoring report (Eurostat 2011):

Investment: "Investment by institutional sectors"

DMC: "Domestic material consumption by material"

CO₂: "CO₂ emissions per inhabitant in the EU and in developing countries"

Energy: "Gross inland energy consumption by fuel"

Waste: "Municipal waste treatment, by type of treatment method"

Bird: "Common bird index"

Transport (EU-15): "Modal split of passenger transport"

Ozone: "Urban population exposure to air pollution by ozone"

Source: Eurostat, EU-27

Table 2.8: Details of the assessment results of the category 4 assessment methods tested on eight indicators

	Belgium	Eurostat	Germany	Netherlands	Switzerland	United Kingdom
Total investment	😊 (0)	☁️ (-0.5)	⚡️ (-1)	➡️ (0)	❌ (-1)	❌ (-1)
Domestic Material Consumption (DMC)	😊 (0)	☀️ (0.5)	☀️ (1)	➡️ (0)	❌ (-1)	✅ (1)
CO ₂ emissions	😊 (0)	☀️ (0.5)	☀️ (1)	↕️ (1)	✅ (1)	✅ (1)
Energy consumption	😊 (0)	☁️ (-0.5)	⚡️ (-1)	↕️ (-1)	❌ (-1)	❌ (-1)
Waste generated	😊 (0)	☁️ (-0.5)	⚡️ (-1)	↕️ (-1)	❌ (-1)	❌ (-1)
Common Bird Index	😊 (0)	☁️ (-0.5)	⚡️ (-1)	➡️ (0)	🌀 (0)	❌ (-1)
Modal-split of transport (EU-15)	😊 (1)	☀️ (1)	☀️ (1)	↕️ (1)	✅ (1)	✅ (1)
Air pollution by ozone	😞 (-1)	⚡️ (-1)	⚡️ (-1)	➡️ (0)	❌ (-1)	❌ (-1)

The highest variance can be observed for the indicator ‘Domestic Material Consumption’ (0.58) followed by the ‘Total investments’ and the ‘Common Bird Index’ indicators (both 0.24). An explanation for these important dispersions is the general shape of the indicators in question. The values of the indicators ‘Total investments’ and ‘Common Bird Index’ vary greatly over time. As for the indicator ‘Domestic Material Consumption’, there is a break in the curve at the last year of the period analysed. The different methods react in different manners to uneven evolutions of the indicators because they do not take into account the same values of the indicators (see Figure 2.15 and Table 2.5 describing the very different results obtained by the Swiss and UK assessment methods with the indicator ‘Domestic Material Consumption’ for an example of this).

2.3 Synthesis

2.3.1 Assessment method considerations

2.3.1.1 The methods are not mutually exclusive

As already mentioned at the beginning of section 2.2, the methods presented in the sections 2.2.1 to 2.2.4 are not mutually exclusive. In Belgium, for example, both category 1 and category 4 methods are used in parallel for some indicators (those for which a target value and a target year are available).

It is also conceivable to apply methods from different categories to different indicators, e.g. category 1 method for indicators with a target value and a target and category 4 methods for indicators with a desired direction only.

The most important point is to communicate clearly and in a transparent manner how the indicators are assessed (see Chapter 3).

2.3.1.2 Indicators not evolving in a clear manner

Not all indicators are evolving in a clear manner, e.g. constant increase or decrease over time. Some indicators have an uneven shape, with successive ups and downs, or a U shape, etc. This complicates the assessment. Lacking data (e.g. a new data point every five to ten years for a period analysed that is 10 or 20 years long) also complicates the assessment.

The assessment method needs to be adapted case by case in these situations, and these exceptions to the standard method have to be duly documented and communicated (see Chapter 3).

It is also possible that a trend change occurs in the period analysed for some indicators. Such changes can be highlighted by using both a short and long-term assessment. In the UK ‘Sustainable Development indicators in your pocket’ publications (see e.g. DEFRA 2010), an assessment is made against a baseline of 1990, 2003 (1999 instead of 2003 until the 2008 edition) and for the penultimate year in order to account for changing trends (three periods analysed).

2.3.1.3 Use of averages

In the trend assessment methods in use (category 4 methods), averages are often calculated in order to minimize the impact of an erratic value. The methods based on a target value and target year (category 1 methods) do not use averages, although the risk that a specific value of the indicator influences the result of the assessment is the same. Not using averages for the calculation of the observed evolution implies that the result of the assessment may change more easily from year to year (instability of the assessment results over time). Depending on the target audience, this potential frequent change may be seen as a good thing helping to capture attention on the indicators, or on the contrary the user could expect a more ‘robust’ assessment that changes only if the general evolution of the indicator is changing.

2.3.1.4 Historical path of the indicator

All the methods currently in use except one do not take into account the historical path of the indicator. Only the methods for trend assessment used by the Netherlands takes all values of the indicator into account (linear regression).

For category 1 methods and category 2 methods (target value available), the possibility of taking into account the historical path of the indicator has not been explored yet. It could consist in fitting a simple statistical model to the data and extrapolate to the target year. The statistical model could be a simple linear regression (such as the trend assessment method of the Netherlands) or a parsimonious curve (if the latter provides a better fit to the data).

2.3.1.5 Use of scenarios

For some indicators, scenarios (expert or model-based projections) are published that may be more or less favourable in terms of assessment than an assessment based on a simple statistical projection (see section 2.2.1.2). Such projections may contain additional information or assumptions that are not taken into account in a simple statistical projection based on past values of the indicator only. The question remains open as to whether such projections, when available, should be used in place or parallel to simple statistical projections. Additional reflections should also be made as to the type of projections that can be used by official statistics (e.g. international, national, or academic projections), and as to how to integrate updates of the projections in the assessment. Finally, if projections were to be used, the rationale behind using such projections would need to be clearly communicated.

2.3.2 Linear/non-linear description of the observed and desired evolution

A linear description of the observed or desired evolution of an indicator implies a constant amount of change (constant slope), with a changing rate of change. This is illustrated in Figure 2.18a.

A non-linear description implies a constant rate of change, with a changing amount of change. This is illustrated in Figure 2.18b.

In practice, both linear and non-linear methods are used. The choice of the one or the other method is rather pragmatic and does not rely on the shapes and expected progress of the individual indicators (the methods currently in use are usually meant to be applied uniformly on all indicators).

2.3.2.1 Category 1 methods (target value and target year available)

It is interesting to note that the assessment based on a target path is not fundamentally different from the one based on a projection. Actually, if both the target path and the extrapolation were calculated in a linear manner (straight line, constant slope) instead of an exponential manner (curve, constant rate of change e.g. AAGR), the result of the assessment would be the same (provided that the values of the indicator taken

into account in the calculation and the assessment classes are the same). The difference between the linear and non-linear calculations of the target path and extrapolation is illustrated in Figures 2.19 and 2.20. In both target path and extrapolation methods, the assessment considers the ratio of two values:

$$\text{Target path methods: } \frac{y_t - x_{t_0}}{x_t - x_{t_0}}$$

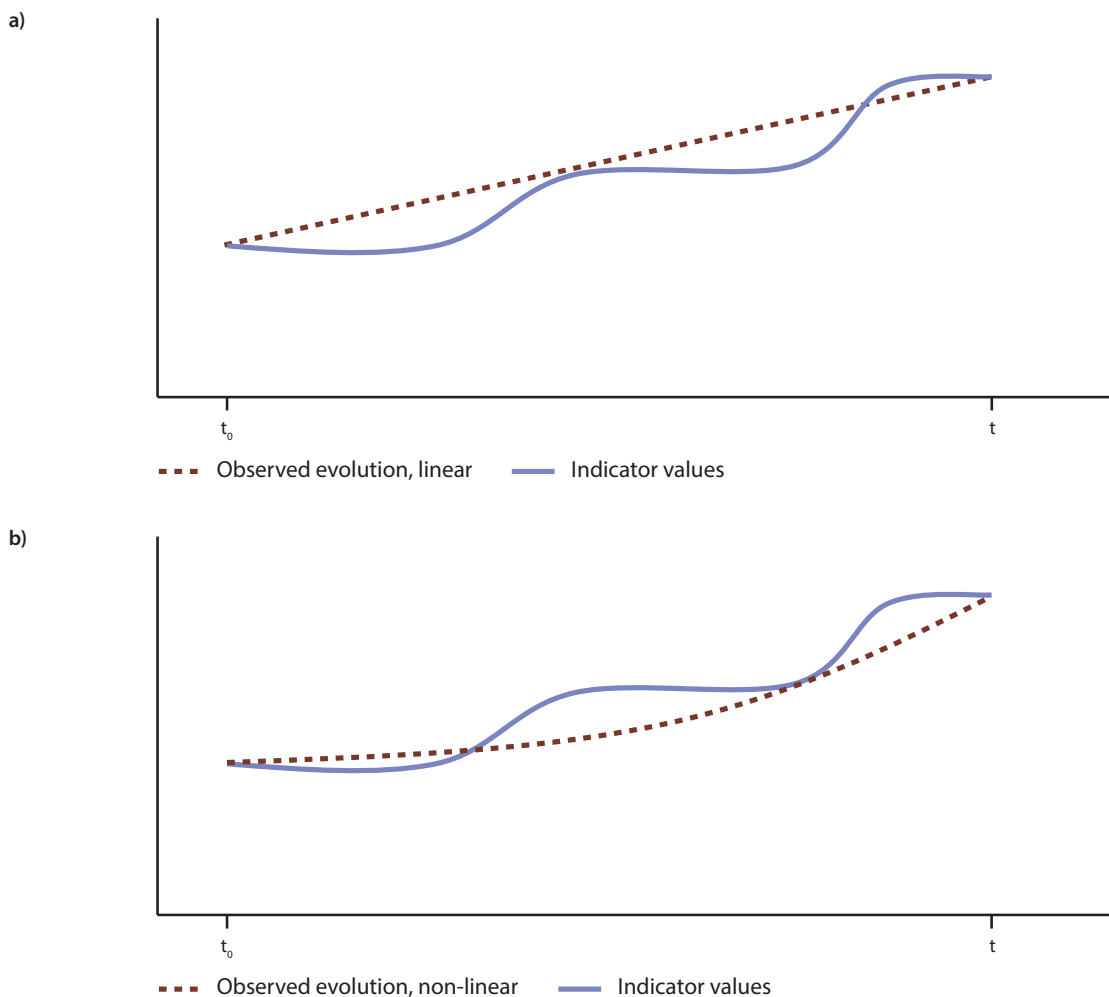
(distance between the reference value x_{t_0} and the last available value y_t over distance between the reference value x_{t_0} and the target path value x_t).

In the Figures 2.19 and 2.20, $(y_t - x_{t_0})$ is represented by the distance A (dark green), and $(x_t - x_{t_0})$ by the distance B (light green). The ratio for target path method is therefore A/B.

$$\text{Extrapolation methods: } \frac{y_{t_1} - x_{t_0}}{x_{t_1} - x_{t_0}}$$

(distance between the reference value x_{t_0} and the projected value y_{t_1} over distance between the reference value x_{t_0} and the target value x_{t_1}).

Figure 2.18: Description of the observed evolution in a linear manner (a) and in a non-linear manner (b)



In the Figures 2.19 and 2.20, $(y_{t1} - x_{t0})$ is represented by the distance C (dark green) and $(x_{t1} - x_{t0})$ by the distance D (light green). The ratio for the extrapolation method is therefore C/D. If both, the target path and extrapolation, are calculated in a linear manner, then $A/B = C/D$ (homothety). This is obviously not true if either the target path or the extrapolation or both are calculated in an exponential manner.

2.3.2.2 Category 4 methods (trend assessment)

For the category 4 methods considering only the first and last available values of the indicators in the period analysed (such as

the ones used by Belgium, Eurostat, Germany, Switzerland and the UK), the choice of the method for calculating the **observed evolution**, on the basis of which the **thresholds** delimitating the assessment classes are set, has an impact on the results. Indeed, if an **AAGR** is chosen, the assessment results will easily be either positive or negative for the first values of the indicator after the **reference value**. After a certain amount of years, however, it will be more and more difficult for the indicator values to go beyond the range of ‘neutral’ assessment results, even if the indicator actually develops in one given direction. It actually requires the indicators to grow or decrease in an exponential manner, which is a strong assumption about its evolution.

Figure 2.19: Representation of the target path and extrapolation calculated with an AAGR (exponential)

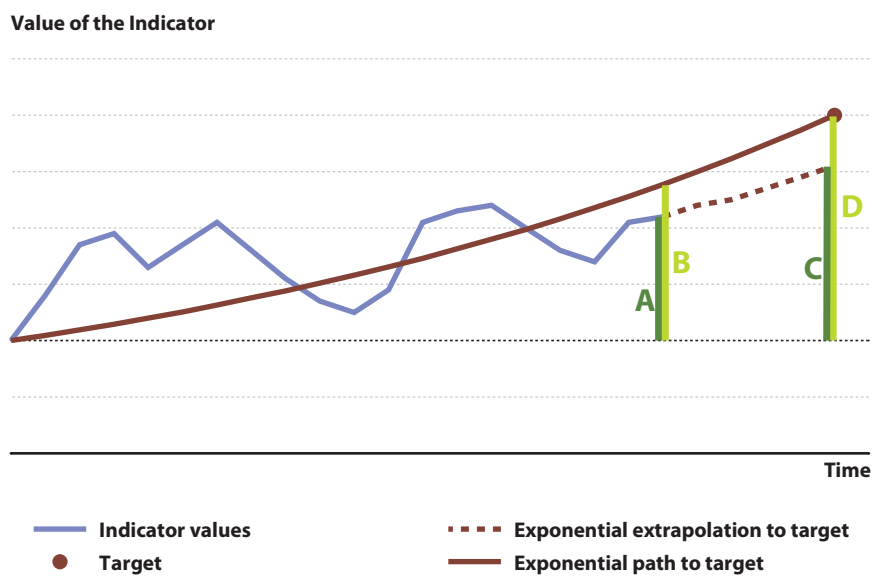
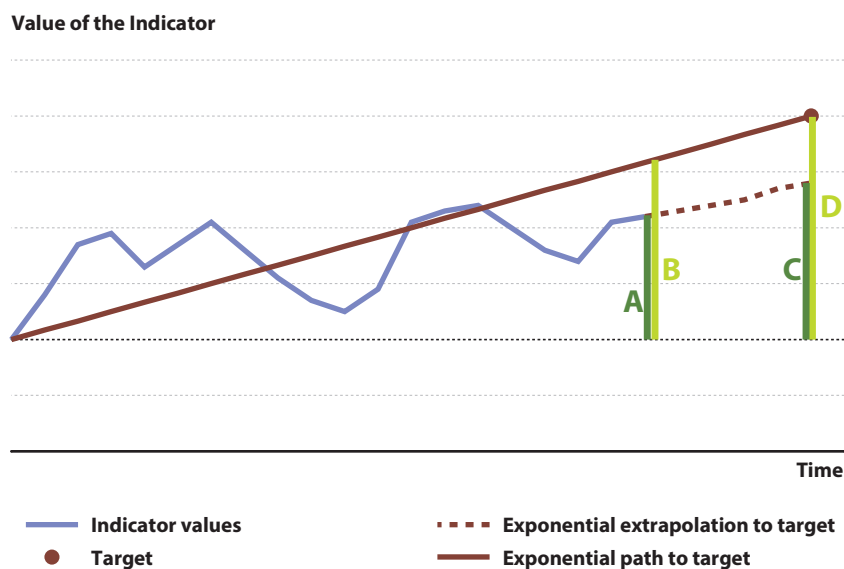


Figure 2.20: Representation of the target path and extrapolation calculated with a constant slope (linear)



If the observed evolution is calculated as a percentage of change (UK and Swiss methods), the opposite is true: the assessment results will easily stay in the 'neutral' range for the first values of the indicator after the reference value. After a certain amount of years, it will be always easier for the indicator values to reach the constant threshold set by a given percentage of change, provided that the indicator actually evolves in one given direction.

The discrepancies between the AAGR and percentage of change methods are illustrated in Figure 2.21.

The choice between a linear and a non-linear description of the evolution of the indicator (observed evolution) is also crucial when using regression models taking into account all the values of the indicator, such as the one used by the Netherlands (see section 2.2.4.2). The significance of the slope is used as a threshold between the assessment classes. One should be aware, however, that the model may not detect significant but non-linear evolution (e.g. an exponential evolution).

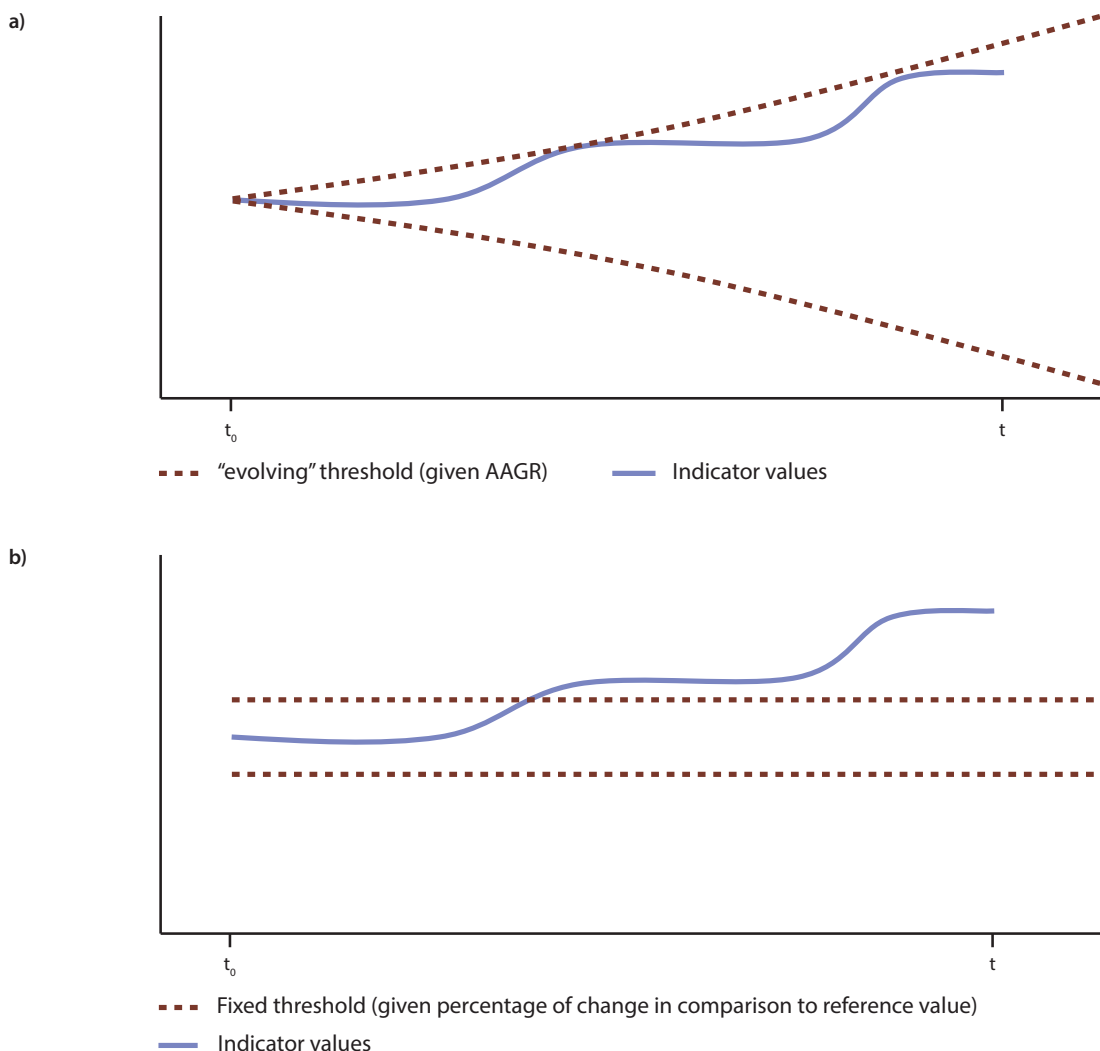
2.3.3 Significance and relevance of the observed evolution

Two elements should be taken into account in order to determine whether or not the **observed evolution** of the indicator corresponds to its **desired evolution** (see also Hulliger, Lussmann 2010):

a) Statistical significance of a change

All or most of the indicators rely on statistical data bound with a certain error. If the change of the indicator value over time remains within the margin of error of the values, the change should not be considered as significant. It is advisable to contact the data producers and ask for advice on the significance of the evolution of the indicator over the **period analysed**. In the UK for example, the data producers are asked to review the assessment and consider statistical significance.

Figure 2.21: Discrepancies between methods using a) AAGR and b) percentage of change for calculating the observed evolution and setting the thresholds



b) Relevance of a change

The change of the indicator value over time might be significant (be higher than the margins of error), but still not mean a real change of the observed phenomenon: it might simply lay within the normal variability of the observed phenomenon.

For example, CO₂ emissions in countries with a temperate climate may increase or decrease on a yearly basis essentially depending on winter weather conditions: particularly harsh winters will induce an increase in CO₂ emissions, due to the increase of fossil fuel consumption used for heating, and particularly mild winters will induce a decrease in CO₂ emissions. A relevant change is observed if the CO₂ emissions increase or decrease more than this 'normal' variability caused by changing yearly winter atmospheric conditions. Such a 'relevant' change can be caused e.g. by an economic recession or by energy saving measures.

Unfortunately, it is even more difficult to obtain information on change considered as relevant from the experts producing the data than on the statistical significance.

It is often impossible to distinguish the two sources of variability in practice (i.e. variance of the data and variability of the phenomenon) (see also Hulliger, Lussmann 2010). From the experience made with SDIs, it is even difficult to obtain any information on the significance and/or relevance of change over time from the data producers.

Instead, several of the **indicator-based assessment** methods in use integrate a certain 'margin of error' in a very rough manner, either by defining a minimum rate or amount of change as a **threshold** for an evolution to be considered significant and relevant, or by applying general significance tests on the time series. These tests are however neither based on the standard deviation of the data used in the indicators (if survey data) nor integrate information on the variance of the data that would be provided by the data producers.

How the topic of non-significant or non-relevant changes is taken into account in the methods currently in use is detailed below for category 1 and category 4 methods.

2.3.3.1 Category 1 methods (target value and target year available)

Regardless of how the **observed evolution** is compared to the **desired evolution** (comparison with a linear **target path**, a non-linear **target path**, or non-linear **projection to target**), it is interesting to look at how the **thresholds** for a positive, neutral and negative assessment are defined: this tells how optimistic or pessimistic the assessments are.

The Belgian method (linear target path) attributes the **indicators** a positive assessment if the last available value of the indicator amounts to 80 % or more of the target path value, which allows a 20 % deviation from the target path. The Swiss method (also linear target path) only attributes the indicator a positive assessment if the last available value amounts to 97 % of the

target path value, which means that only 3 % deviation from the target path is tolerated.

One could say that the Belgian method is more optimistic, allowing 20 % uncertainty in favour of a positive assessment. On the other hand, the area for a neutral assessment (instead of a negative assessment) is larger in the Swiss method: as long as the indicator evolves in the **desired direction**, the assessment is neutral. The Belgian method attributes a negative assessment as soon as the ratio is lower than 50 %, even if the indicator evolves in the desired direction, which does not give much room to the neutral area (between 50 and 80 % of the target path value only). It is the only method in this category that attributes a negative assessment even if the observed direction corresponds to the desired direction. See Figure 2.4 for a visual representation of the above.

The Eurostat method (non-linear target path) attributes a positive assessment only if the rate of change of the indicator (AAGR calculated on the basis of the **reference value** and the last available value) amounts to 100 % or more of the desired rate of change (AAGR calculated on the basis of the reference value and the **target value**). This can be considered as rather strict and pessimistic. However, the indicator is attributed to a 'moderately positive' class if the observed rate of change lies between 80 and 100 % of the desired rate of change. As for the Swiss method, a negative class is attributed only if the indicator evolves in the direction opposite to the desired direction (a 'moderately negative' class is attributed if the indicator evolves in the desired direction but at a rate of change lower than 80 % of the desired rate of change).

The German method (projection to target) attributes a positive assessment if the projected value of the indicator at **target year** amounts to 95 % of the target value.

2.3.3.2 Category 4 methods (trend assessment)

2.3.3.2.1 Rate or amount of change as a threshold

The Belgian trend assessment method (see description in section 2.2.4.1.1) delimitates the neutral class with a **threshold** of a +/- 1 % AAGR for the **period analysed**: as long as the AAGR of the indicator in the period analysed (**observed evolution**) does not exceed 1 % or -1 %, the indicator is allocated to the 'neutral' class. One can interpret that an AAGR of the indicator not exceeding 1 % in absolute terms in the period analysed means that the indicator is not considered as having evolved significantly nor in a relevant manner in this period.

The United Kingdom and Switzerland also use a 'neutral' class, delimited by a threshold of a 3 % change in absolute terms over the period analysed: as long as the percentage of change of the indicator in the period analysed (**observed evolution**) does not exceed 3 % in absolute terms, the indicator is allocated to the 'neutral' class, meaning that it is not considered as having evolved significantly and/or in a relevant manner in this period.

Eurostat uses the same thresholds as Belgium (AAGR of 1 % in absolute terms) but adds a further threshold at an AAGR of 0 %. The consequence is that there is no unique ‘neutral’ assessment class: an AAGR between 0 % and 1 % in the *desired direction* means that the observed evolution is considered as ‘moderately positive’, and an AAGR between 0 % and 1 % but opposite to the desired direction means that the observed evolution is considered as ‘moderately negative’. It is probably more difficult for the user to interpret these ‘moderate’ classes as depicting not significant or not relevant evolutions of the indicators.

2.3.3.2.2 Use of significance tests

The Netherlands test the slope of the regression carried out on the *indicator* values to know whether it is significantly different from zero. If the slope is not significantly different from zero, the indicator is attributed to the ‘neutral’ class.

Germany does not make use of an actual ‘neutral’ class. There is only one *threshold* at an AAGR of 0 %: if the sign of the AAGR of the indicator corresponds to the *desired direction*, it is attributed to the ‘positive’ class, and if the sign of the AAGR does not correspond to the desired direction, it is attributed to the ‘negative’ class. The significance of the evolution is, however, tested by means of a Spearman’s rank correlation coefficient. The result of the test is indicated parallel to the ‘positive’ or ‘negative’ class.

2.3.4 Which method should be used?

The choice between the various assessment categories is quite clear: as mentioned at the beginning of this chapter, it is possible to carry out category 1 methods only if a *target value* and

a *target year* is available, and category 2 assessment methods only if a target value is available. Since the definition of target values and target years are in practice not under the responsibility of those carrying out *indicator-based assessment*, but rather the task of policy-makers, this is a decisive piece of information in the choice of an assessment category.

As it is clear from the sections 2.2.1 to 2.2.4 above, the different methods within one assessment category summarize the *indicators* in a different manner depending on the values of the indicator taken into account by the method, but also on the shape of the indicator. One could consider carrying out a simulation to test how the methods react to different shapes and particular values of indicators, for example. This would require a deeper analysis and is out of the scope of this handbook.

The most important advice from this chapter is therefore not about the choice of a precise assessment method, but of a quite different nature: to avoid relying solely on the assessment method chosen. Indeed, it is very important to consider the overall shape of the indicator and the context in which it is being assessed. This also means that one should not only focus on the assessment results in the form of pictograms (symbols) (see Chapter 3), but that the indicator-based assessment process should include an analysis of the situation.

Indicator-based assessment is a powerful tool for synthesizing information. It simplifies the message given by indicators. However, delivering a *simple* message should not be mixed up with delivering a *simplistic* message. Carrying out indicator-based assessment implies being aware of the complexity that hides behind the simple message and taking account of it for communication purposes (see Chapter 3).

**Communication of
indicator-based assessment**

3

3.1 Context

Chapter 2 describes a variety of assessment methods that are applied to convey an overall state of affairs with respect to sustainable development indicators (SDIs). Three categories of assessment methods are used in practice: methods based on a target value and a target year (category 1 methods), methods based on a target value only (category 2 methods) and methods based on a desired direction (category 4 methods, also called ‘trend assessment’ methods). Subsequently, these assessment methods can be communicated at the level of individual indicators, groups of indicators and in an aggregated manner (see Box 3.1). It is of great importance to provide transparency in the assessment process and its results to comply with the quality principles of official statistics, notably Principles 6 (impartiality and objectivity) and 15 (accessibility and clarity) of the European Statistics Code of Practice (ESSC 2011). The user needs to receive a clear and unambiguous message. A major challenge is therefore to simultaneously provide transparency with this clear message.

Box 3.1: Communication levels of the assessment results

I — Communication at *individual indicator level* means that the result of the assessment is communicated at a level where the indicator is presented (e.g. next to graph).

G — Communication at *the level of groups of indicators* means that the result of the assessment for an individual indicator is presented in combination with several indicators (e.g. in lists or tables). The communication of several indicators in lists or tables is also commonly referred to as ‘**dashboard**’. This understanding of the term ‘dashboard’ is in line with the Stiglitz report (Stiglitz, Sen, Fitoussi 2009, see e.g. paragraphs 131 and 132 in section 2.1).

A — Communication in an *aggregated manner* means that the result of the assessment is communicated for a number of indicators collectively (e.g. per theme in circle diagrams) which, *if it was presented alone*, would not allow the user to know which indicator contributed to the aggregated information (theme, dimension) nor to what extent. This communication level is also commonly referred to as ‘**dashboard**’. In order to distinguish the dashboard at an aggregated level from a dashboard at a grouped level they will be further called ‘**aggregated dashboards**’ in this handbook. There are two sub-categories for this communication level:

- **Simple, non-interactive aggregated dashboard** in printed publications (a typical example is the summary of assessment results in a pie chart in the UK brochure presenting their sustainable development indicators (DEFRA 2010).
- **Interactive aggregated dashboard** — the result of the assessment is communicated for a number of indicators collectively in an interactive manner (online), which allows a user to trace how each indicator contributed to the aggregated information (e.g. the Swiss and Dutch aggregated dashboards in section 3.3.3).

This chapter focuses on the communication of the process and of the results of *indicator-based assessment*. The main objective of this chapter is to provide a number of essential elements (criteria) that a newcomer in the field of indicator-based assessment should consider in order to communicate the assessment process and results effectively. The chapter will compare approaches used by the Expert Group members regarding the indicator-based assessment communication at the level of individual indicators, groups of indicators and in an aggregated manner. There is no intention to point out especially good or bad examples across the

Expert Group countries. The objective is rather to analyse some examples in a systematic way, and to illustrate in a concrete manner the main elements for a good (transparent) communication of the indicator-based assessment processes and results.

The chapter starts with the overview of various communication forms (printed versus online) employed by the Expert Group members. After that the communication of the process and results of indicator-based assessment is presented and discussed; first at the level of individual indicators, then at the level of groups of indicators (dashboards), and finally at an aggregated level (aggregated dashboards). The chapter ends with a discussion on the symbols used to communicate the results of the indicator-based assessment.

3.2 Communication forms and target audiences

SDIs are communicated in different ways according to the target audience. The amount of information provided about the process and result of *indicator-based assessment* depends on the communication form chosen and the target group. Table 3.1 provides an overview of the communication forms applied by the Expert Group members.

Table 3.1: Communication forms used in the Expert Group countries

Countries/ communication forms	Printed publication or PDF publication available online		Online indicators	
	Report	Brochure	Dashboard/ group of indicators	Interactive aggregated dashboard
Belgium	V	—	V	—
Eurostat	V	—	V	—
Germany	—	V	—	—
Netherlands	V	V	V	V
Switzerland	V	V	V	V
United Kingdom	—	V	V	—

V: available
—: not available

Printed publication — report

- Comprehensive character — it contains a detailed description of the theoretical background, *indicators*, assessment methods.
- The target audience is interested in detailed information.

Printed publication — brochure

- Incomprehensive character — it is intended to present only the core message, it is distilled from a mass of data. Considering the limited size of a publication, most of the details cannot be presented. It often provides an overall state of

affairs. A reference (link) to detailed description is therefore indispensable in order to comply with the principle of clarity of official statistics.

- The target audience is interested in getting an overview of the situation/progress.

Online indicators — dashboards/individual indicators

- Synthetic character but quick access to the underlying data.
- The target audience is interested in getting the most relevant information on the issue of sustainable development.

Online indicators — interactive aggregated dashboards

- Inherent to this communication form is that information is provided in a layered manner. The use of an interactive aggregated dashboard allows the user to get an overview of the situation (aggregated information) while at the same time providing access to detailed information (at indicators' individual level). Also references to additional information concerning the theoretical background, assessment methods etc must be provided.
- The target audience is interested in getting an overview of the situation (progress) and in having access to the underlying data.

Dependent on the communication form and the target group, diverse choices are made with regard to the amount of information provided on indicator-based assessment. For instance, a brochure is intended for the general public and due to its small format does not need to contain all methodological details. However, a reference to a detailed methodology is indispensable with such a brochure in order to ensure transparency.

The communication of the process and results of indicator-based assessment of individual indicators, groups of indicators (dashboards) and in an aggregated manner (aggregated dashboards) will be discussed respectively in the following sections. A distinction is made between printed publications and online indicators. For a complete overview of the publications and websites that were used for the analysis presented in the following sections, see Annex 2.

3.3 Communication of the process and results of indicator-based assessment

3.3.1 Communication at the level of individual indicators

The assessment of an individual indicator is performed by comparing the observed evolution with the desired evolution. The objective of the indicator-based assessment is to synthesise

the message of the indicators (see the Introduction and Chapter 1). By presenting the assessment results with symbols the information contained in the results can be condensed and easily accessed.

As already mentioned in Chapter 2, three categories of assessment methods are used in practice:

- Methods based on a **target value and a target year** (category 1 methods) — section 3.3.1.1.
- Methods based a **target value only** (category 2 methods) — section 3.3.1.2.
- Methods based on a **desired direction** (category 4 methods, also called 'trend assessment' methods) — section 3.3.1.3.

The discussion of each category of assessment method (sections 3.3.1.1, 3.3.1.2 and 3.3.1.3) consists of the following headings:

Essential elements to be communicated

As described in Chapter 2, a number of essential elements needs to be defined for each assessment method, so that a method can be implemented. Communicating these elements contributes to the transparency of the assessment process. They can therefore be considered as criteria for a good communication of the assessment process.

Experiences

The communication of the essential elements will be discussed and supported with individual examples from Expert Group countries. It should be noted that examples used in this part (charts, graphics) are presented outside their context. In some cases, a piece of information that is important for the transparent communication of the assessment process and results is not communicated on the graph of the indicator but somewhere else in the publication, dependent on the communication form.

The section 3.3.1.4 provides a summary and a discussion on the communication of the assessment process and results at the level of individual indicators.

3.3.1.1 Communication of category 1 methods (target value and target year available)

Some indicators are tracked according to time-bound target values. Such target values have to be attained in a certain period of time. By using target values accompanied by a target year it is possible to monitor the progress of an indicator towards a goal over time:

- Positive: changes in the **desired direction** and at the required pace (slope or rate of change) needed to attain the target value.
- Negative: change in the undesired direction needed to attain the target value.
- Neutral: no change.

Eurostat and Germany distinguish four classes: positive, moderately positive (progress in the right direction but too slow),

moderately negative (slow progress in the wrong direction) and negative.

Essential elements to be communicated for category 1 methods

- Target value and target year
- Base year
- Reference value
- Calculation method for the target path or the projection to target
- Actual value(s) of the indicator taken into account for the comparison with the target path/ target value
- Methods used for comparing the observed evolution with the desired evolution
- Result of the assessment (symbols)
- Thresholds used for the assessment
- Source of the target value.

Experiences

Figure 3.1 shows the graph of the indicator 'Greenhouse gas emissions' as presented in the German brochure 'Sustainable Development in Germany: Indicator Report 2012' (DESTATIS 2012). Three target values are presented (2010, 2020 and 2050). The annex of the brochure specifies that the assessment is carried out on the basis of the first target value presented (in this case 2010). The preface indicates the source of the targets (defined by the German Federal Government in the context of the National Strategy for Sustainable development). In this example, a base year is indicated in the graph, which is the base year used for the calculation of the index ⁽²⁰⁾. This base year should not be confounded with the base year defined in Chapter 2, section Terminology. The observed evolution of the indicator is calculated by means of an Average Annual Growth Rate (AAGR) over the last five years. This means that it is based on two values of the indicator: the last available value and the value five years before. The value of the indicator five years before the last available value is projected by means of the AAGR to the target year. The projected value at target year is then compared to the target value. This is partly explained in the annex of the German brochure (the year at which the projection begins is not

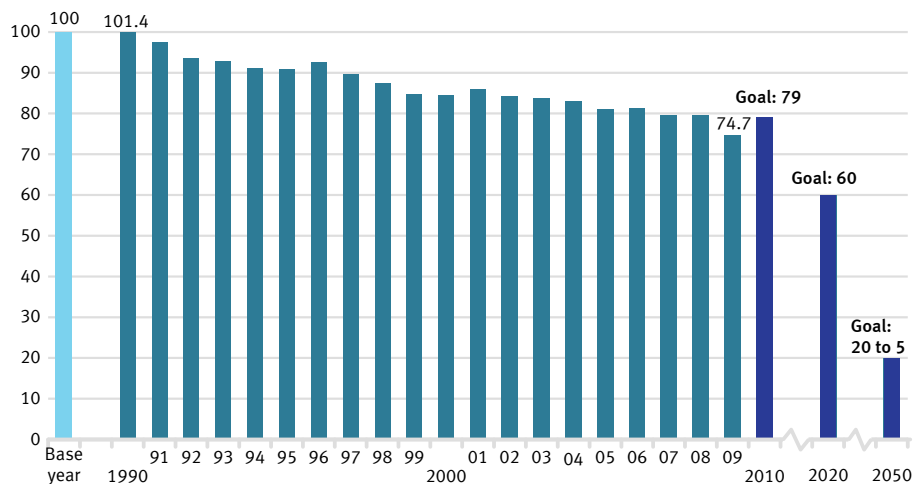
Figure 3.1: Graphical representation of an indicator assessed with a category 1 method — example in Germany

Climate protection



Reducing greenhouse gases

Greenhouse gas emissions (six Kyoto gases) in CO₂ equivalents Base year = 100



Source: Federal Environment Agency

Source: DESTATIS 2012 (brochure)

NB: the base year indicated in this graph does not correspond to the base year defined in this handbook. See the definition of 'base year' used in this handbook in section 2.1.1 in Chapter 2 and the text below this figure for more information.

⁽²⁰⁾ The base year for the calculation of the index stems from the Kyoto Protocol and is a construct between the data for CO₂, CH₄ and N₂O — for which the Kyoto base year is 1990 — and the data for HFC, PFC and SF₆ — for which the Kyoto base year is 1995.

indicated). It should be noted that the user cannot discern the period analysed before reading the annex. The graph presents values starting in 1990 and the period being analysed is not highlighted. The user will probably deduce that the assessment concerns the period 1990–2009. Hence a reference to the methodology in order to guide the user is essential in such cases.

The thresholds that delimit the possible assessment results are also indicated in the annex. Additionally the weather symbols present a result of assessment of the progress towards this target. The symbols are communicated at the level of an individual indicator (above the graph).

Figure 3.2 shows the ‘Total employment rate, EU-27’ indicator as it is presented in the report ‘Sustainable Development in the European Union’ (Eurostat 2011).

The target value, the reference value at base year (2000) and the last available value of the indicator are presented. The graph may not be completely self-evident concerning these values e.g. the reference value is not easily discerned. However, the period analysed is indicated in the introduction of the report, together with the calculation method for the target path, the values of the indicator used in the comparison, the methods for comparing the observed and desired direction, and the thresholds delimiting the possible assessment result (p. 41–44 of the report). The source of the target value (EU Sustainable Development Strategy) is also indicated in the introduction of the report. The observed and desired evolutions are summarized in the information given on the right side on the graph (Average Annual Growth Rate and required Average Annual Growth Rate). The indication of the value on the target path at the year of the last available value (value 70.7 % on the graph) and of the distance to target path in 2010 (–2.1 percentage points, indicated on the

right side of the graph) give an information about how far the indicator is from the target path. However these two pieces of information are not taken into account in the assessment process, which consists of calculating the ratio of the (observed) AAGR and of the desired AAGR. The result of the assessment is communicated with a weather symbol. It should be noted that the same type of symbol is used for the communication of the result of category 4 (trend) assessment.

3.3.1.2 Communication of category 2 methods (target value available, no target year defined)

For some indicators, the target value has been defined without specifying the target year.

Essential elements to be communicated for category 2 methods

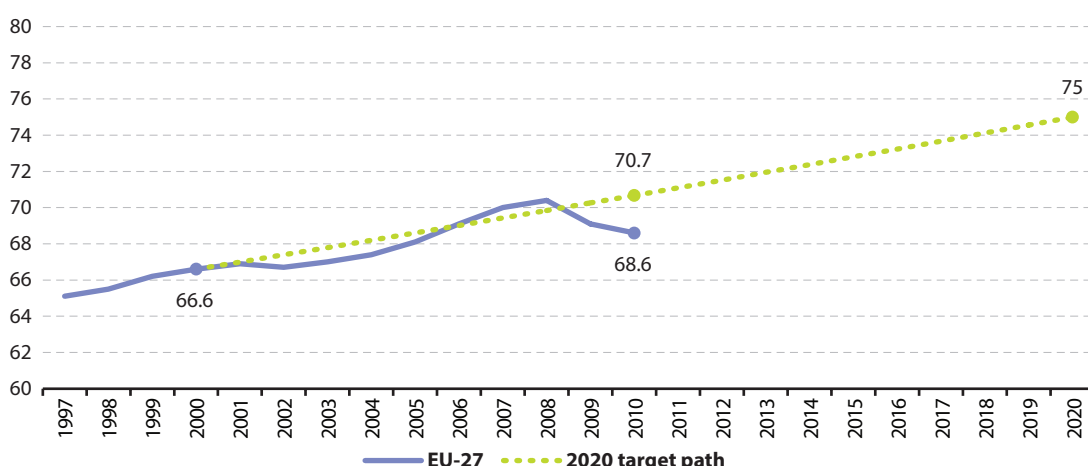
- Target value
- Value(s) of the indicator taken into account for comparing with the target value
- Method used for comparing the value(s) of the indicator with the target value
- Result of the assessment (symbols)
- Thresholds used for the assessment
- Source of the target value.

Experiences

Figure 3.3 shows the indicator ‘Land used for organic farming’ as it is presented in the German brochure ‘Sustainable

Figure 3.2: Graphical representation of an indicator assessed with a category 1 method — example Eurostat

Figure 1.14: Total employment rate, EU-27 (% of age group 20-64 years)



Period evaluated:
2000-2010
Average annual
growth rate:
+0.3 %
Required annual
growth rate:
+0.6 %
Distance to target
path in 2010:
-2.1 percentage
points

Source: Eurostat 2011 (report)

Development in Germany: Indicator Report 2012' (DESTATIS 2012). The **target value** is indicated on the graph. The origin of the target is mentioned in the preface of the brochure. Neither the value(s) of the indicator taken into account for comparison with the target value nor the method used for the comparison are indicated in the brochure for this specific type of assessment. The **thresholds** delimiting the assessment results are not indicated. The assessment based on a target value without a target year can be considered as an exception to the 'standard' assessment for the German SDIs, which is based on a target value with a target year. The methodology for this 'exception' is not documented (see section 3.3.1.4. for discussion on communication of exceptions). The assessment result is communicated at the level of an individual indicator (above the graph).

3.3.1.3 Communication of category 4 methods (trend assessment)

The evaluation of trends is based on the evolution of an **indicator** in a defined period (**period analysed**). The evaluation considers the direction of change:

- Positive: change in the desired direction
- Negative: change in the undesired direction
- Neutral: no change.

For more details on trend assessment see Chapter 2, section 2.2.4.

In order to communicate the trend assessment in a transparent manner the following elements should be communicated.

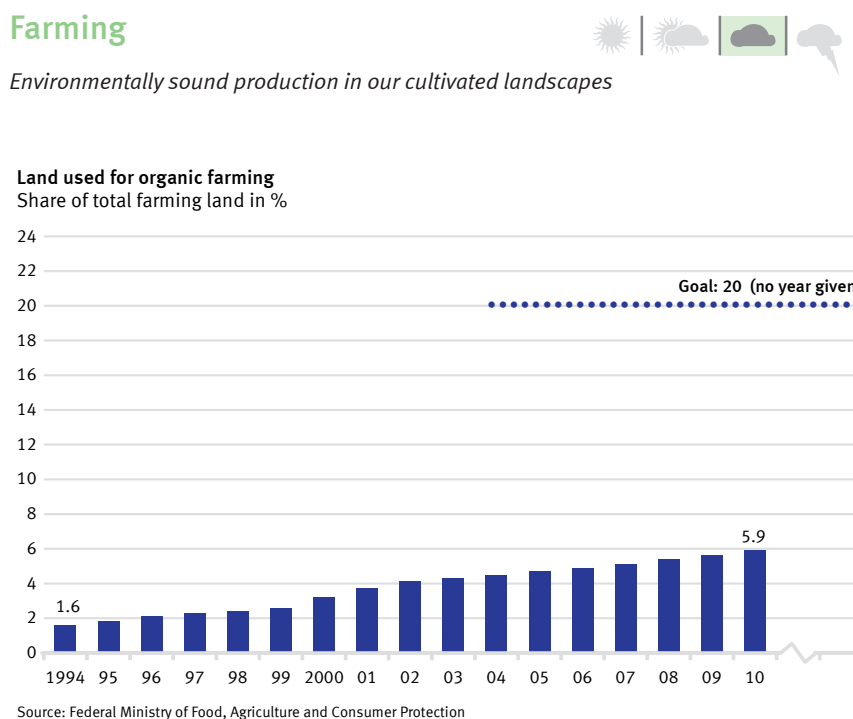
Essential elements to be communicated for category 4 methods

- Period analysed
- Desired direction
- Observed evolution ⁽²¹⁾ (including the methods for its calculation)
- Values of an **indicator** used for calculation of the observed direction
- **Thresholds** delimitating the assessment classes
- Result of the assessment (symbols)
- Source of the desired direction (e.g. the Swiss Sustainable Development Principles) ⁽²²⁾.

Experiences

Figure 3.4 shows the UK **indicator** 'Renewable electricity' as presented in the brochure 'Measuring progress — Sustainable development indicators 2010' (DEFRA 2010). The **desired evolution** is an increase of renewable electricity production. The desired trend is not specifically indicated for any of the indicators presented in the brochure. For the indicator in Figure 3.4

Figure 3.3: Graphical representation of an indicator assessed with a category 2 method — example Germany



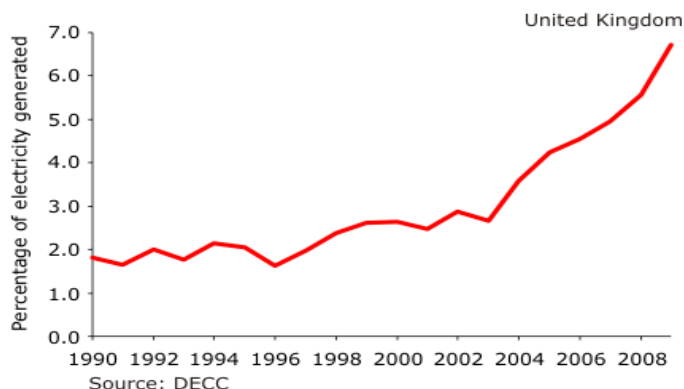
Source: DESTATIS 2012 (brochure)

⁽²¹⁾ In the case of trend assessment methods, the observed evolution simply consists in the observed direction.

⁽²²⁾ Sometimes the desired direction is considered as obvious and is not explicitly indicated (e.g. decrease of unemployment). See section 3.3.1.4 for more details.

Figure 3.4: Graphical representation of an indicator assessed with a category 4 method — example in the UK

Renewable electricity generated as a percentage of total electricity, 1990 to 2009



Renewable electricity since: 1990 2003

Source: DEFRA 2010 (brochure)

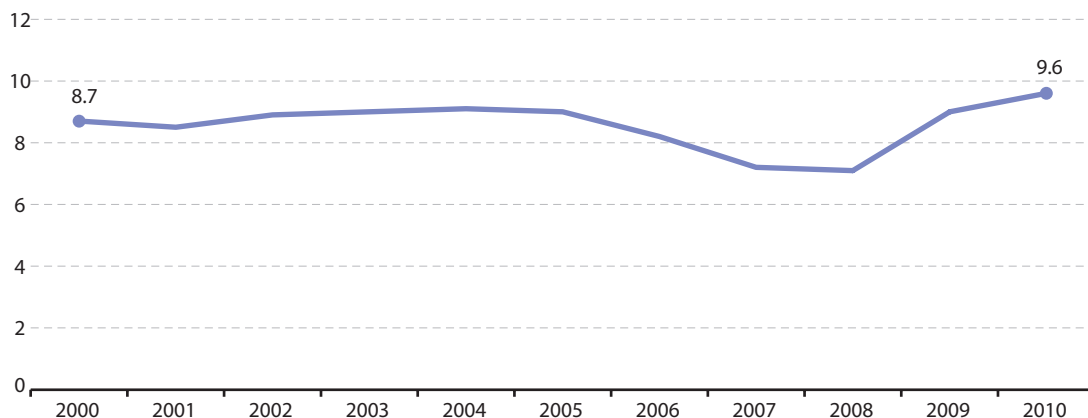
the desired trend can be considered as self-evident, although for other indicators presented in the brochure it may not be so obvious (see section 3.3.1.4. for a discussion on communication of the desired trend). The United Kingdom assesses the evolution of each indicator (in this case renewable electricity) against two reference values: one with 1990 as the base year for long-term assessment and one with 2003 as the base year for short-term assessment. The base years are clearly presented. The period analysed is clearly presented but the exact values for calculating the observed evolution (i.e. the average of the

values 1989–1991 for base year 1990 or the average of the values 2002–2004 for base year 2003) are not specified. This is, however, explained in the annex of the brochure as well as thresholds and the applied methodology. Additionally the symbols ⁽²³⁾ (a set of arithmetic operators in combination with traffic lights colours, see below the graph) are used to communicate the result of the assessment at the level of an individual indicator. In this case the development shows a positive trend towards sustainable development (clear improvement since the base year) as indicated in Figure 3.4.

Figure 3.5: Graphical representation of an indicator assessed with a category 4 method — example in the EU

Figure 1.19: Total unemployment rate, EU-27

(%)



Period evaluated:
2000-2010
Average annual
growth rate:
+1.0 %

Source: Eurostat (online data code: tsdec450)

Source: Eurostat 2011 (report)

⁽²³⁾ An overview of various symbols used across countries and their interpretation is provided in section 3.4 of this chapter.

Figure 3.5 shows the EU indicator ‘Total unemployment rate, EU-27’ as presented in the report ‘Sustainable development in the European Union’ (Eurostat 2011). The desired trend is not explicitly specified. This is not necessarily an issue in this specific example since the desirability of a decreasing unemployment rate can be considered to be self-evident. However, this is not a case for all the indicators presented in the report (see section 3.3.1.4. for a discussion on the communication of the desired direction). Information regarding the period analysed and the result of the assessment is communicated at the individual level. Information about how the calculation is made is communicated in the introduction of the report, together with other information on the assessment methodology such as thresholds. In comparison to Figure 3.4, this graphic is complemented with additional information of exact values at the start and end of the series. These values are used for the calculation of the observed trend.

3.3.1.4 Summary and discussion

Table 3.2 shows whether the essential elements defined for each category of assessment methods are communicated in the print publications and online indicators of the Expert Group countries. The table reveals discrepancies in communicating assessment results. No single country communicates all the essential elements defined. There are still some open issues concerning:

Communication of the period analysed

The time series presented on the graph do not always correspond to the actual period analysed. This should be clearly stated in order to ensure the transparency and clarity of what is being assessed. The period analysed can be indicated e.g. by highlighting it visually on the graph, by stating it next to the graph (see e.g. Figure 3.5) or by providing a direct link (reference) to the underlying information. This is, however, not always the case in practice.

Communication of desired direction

There are cases where the desired direction can be considered to be self-evident (e.g. (un)employment, see Figure 3.2 and Figure 3.5). There are, however, two types of indicators where the desired trend is not always, but should be explicitly communicated:

1. Context-sensitive indicators

There are indicators that can be interpreted in different ways dependent on the context and the phenomenon they represent. For instance, expenditure indicators (such as ‘Expenditure on health’) — it is impossible to say, without knowing

the context, whether an increase or a decrease is desirable. In order to assess these types of indicators extra information is required. Another example of context sensitive indicators is Foreign Direct Investments (FDIs) — increase of FDI can be seen as a contribution to development aid and be considered as a positive development. But an increase of FDI can also be seen as a transfer of the production of developed countries overseas, for instance to developing countries in order to take advantage of lower wages and less stringent regulatory environments, and be considered as negative development.

2. Unfamiliar indicators

Relying on phenomena that only experts are familiar with, for instance ‘Material flow’, ‘Energy trade balance’ or ‘Nitrogen surplus’ — only one familiar with those concepts will be able to indicate the desired trend.

When communicating about the direction and the evolution in general, it is also important to specify which direction is meant (desired vs. observed). For example, in the Belgian report (FPB 2011) only the title ‘Direction’ is provided in table 2 on p. 14 of the French version of the report. At first glance this indicates the observed direction, however, the symbols actually indicate the desired direction.

Communication of exceptions

Another issue is the means whereby the inevitable exceptions to the assessment method are communicated. As mentioned in the synthesis of Chapter 2, in some cases the assessment method needs to be adapted to a particular indicator. This is e.g. the case for the German SDIs lacking a target value and a target year. The general method described in the brochure ‘Sustainable Development in Germany’ (DESTATIS 2012) that is used for most of the German SDIs cannot be applied if the target year is lacking. It is important that this kind of ‘exception’ to the ‘general’ assessment method is communicated. Belgium uses footnotes in the indicator report to explain why the assessment method was changed for the one or the other indicator (FPB 2011).

Source of a target or desired direction

Expert Group countries refer to the source of a target or a desired direction in general terms. For instance, Germany refers in the introduction of their brochure to the National Strategy for Sustainable Development entitled ‘Perspectives for Germany’ (DESTATIS 2012) without specifying the exact postulates. Such information could, however, be useful for the user. Switzerland, for example, provides a precise reference to every target or desired direction for each indicator.

Table 3.2: Overview of the essential elements of the assessment process and results and of their communication

Method	Criteria	Printed publications or PDF-publications available online					Online indicators						
		Belgium ¹⁾	Eurostat	Germany	Netherlands	Switzerland ²⁾	United Kingdom	Belgium	Eurostat ³⁾	Germany ⁴⁾	Netherlands	Switzerland	United Kingdom
Category 1 assessment methods	Target value and target year	I	D	D/I	n/a	D	n/a	D	—	n/a	n/a	D	n/a
	Base year	—	D	—	n/a	D	n/a	D	—	n/a	n/a	D	n/a
	Reference value	—	D	I	n/a	D	n/a	D	—	n/a	n/a	D	n/a
Category 2 assessment methods	Calculation method for the target path or projection to target	—	D/I	I	n/a	I	n/a	I	—	n/a	n/a	I	n/a
	Values of the indicator taken into account for comparison with target path or target value	—	I	—	n/a	I	n/a	D	—	n/a	n/a	I	n/a
	Target value	I	n/a	D	n/a	n/a	n/a	D	—	n/a	n/a	n/a	n/a
Category 3 assessment methods	Values of the indicator taken into account for comparison with target value	—	n/a	—	n/a	n/a	n/a	—	—	n/a	n/a	n/a	n/a
	Period analysed	I	D	—	I	D	D	D	—	n/a	D	D	D
	Desired direction	I	D	I	I	D	—	D	—	n/a	D	D	—
Category 4 assessment methods	Observed direction	—	D	—	D	D	D	D	—	n/a	D	D	—
	Values of the indicator used for calculation of the observed direction	—	I	—	—	I	I	D	—	n/a	D	D	—
	Thresholds used for the assessment	I	I	I ⁵⁾	I	D/I	I	I	—	n/a	I	I	—
Generic criteria	Results of assessment	I	D	D	—	D	D	D	—	n/a	D	D	D
	Method used for comparing the observed evolution with the desired evolution	I	I	I ⁵⁾	D	I	D	I	—	n/a	I	I	—
	Source of the target value or desired direction	I	D	I	I	D	I	D	—	n/a	—	D	—

D: Information is communicated directly on the graph or just next to it (e.g. in the accompanying text, or in a textbox appearing by passing the mouse over for online indicators).
 I: Information is communicated indirectly e.g. in the preface, introduction or annex of a publication for printed/PDF publications, or on an additional webpage, in a downloadable document for online indicators.

—: Not available/unclear

n/a: Not applicable

1) Refers to chapter A2 (category 4 methods) and to chapter A3 (category 1 methods) of the 2011 Belgian report (FPB 2011).

2) "D" stands for information presented either on the graphs or in the tables at the end of each chapter.

3) The Eurostat SD indicators are assessed in the Monitoring report (Eurostat 2011), but not online.

4) Germany publishes the data of the indicators in the form of tables, but not the indicators themselves (graphs, accompanying texts, etc.).

5) Information available only for category 1 methods.

NB: The publications and websites used for this analysis are referenced in Annex 2.

3.3.2 Communication at the level of groups of indicators (dashboards)

As already mentioned in Box 3.1, the communication of assessment results at the grouped level means that assessment results of individual indicators are presented in conjunction with several indicators (in lists or tables). Therefore, this paragraph only provides the graphical examples of the assessment results on a grouped level. Subsequently, the summary table provides an inventory of the communication in this regard by some of the Expert Group countries.

The assessment results are presented in the Swiss 'Sustainable Development report 2012' (FSO 2012) at the level of group of indicators at the end of each thematic chapter (see Figure 3.6). Such a list of indicators consists of the detailed information on the observed and desired evolution, source of the target value or of the desired direction, period analysed and assessment result. Information about how the observed evolution is calculated and how it is compared with the desired evolution or with the target value is provided in the introduction of the report.

The United Kingdom also provides in its brochure an overview of assessment results on a grouped level (see Figure 3.7). The observed and desired evolution, the source of the desired evolution and the period analysed are provided in the annex of the brochure 'Measuring progress — Sustainable development indicators 2010' (DEFRA 2010).

It can be concluded from Table 3.3 that every Expert Group country presents the assessment results at the grouped level in printed publications. However, communication on the grouped level is not broadly employed with websites (only Switzerland and the Netherlands).

3.3.3 Communication at an aggregated level (aggregated dashboards)

Communication in an aggregated manner means that the results of the assessment are communicated for a number of indicators at once (e.g. per theme in circle diagrams) which, if it was presented alone, would not allow the user to know which indicator contributed to the aggregated information and to which extent. Mostly it aggregates the results of indicator assessment per theme or sustainable development dimension thus giving an overall assessment of the situation for this theme or dimension.

The challenge faced by most SDI sets or systems is to synthesize the information provided by a large number of indicators without losing transparency. There are several ways to aggregate data, and two possible avenues are emerging: composite indicators (see comprehensive overview of composite indicators in Kulig, Kolfoort, Hoekstra 2009) and aggregated dashboards (Wachtl, Mayerat Demarne, de Montmollin 2010).

Figure 3.6: Communication of the assessment results at a grouped level — example Switzerland

T 1.2 Documentation on indicator trends

Graph	Targeted trend	Objectives ^a	Period under analysis	Change in %	Observed trend ^b	Assessment	Comments
G 2.1	↗	2b SDS Nr. 7	1992 2007	W: +7.7% M: +8.6%	↗	✓	Synthesis of trends observed for women (W) and men (M) ^c
G 2.2	→	SDS Nr. 7	1991–1993 2007–2009	+20.1%	↗	✗	
G 2.3	↘	2b SDS Nr. 7	1992 2007	+23.1%	↗	✗	
G 2.4	↗	2b SDS Nr. 7	1992/93 2007	+6.2%	↗	✓	
G 2.5	↗	2b/3a SDS Nr. 7	1992/93 2002	+24.4%	↗	✓	

^a Sustainable development principles, challenges of the sustainable development strategy (SDS) or quantified and dated objectives.

^b The trend is interpreted as relevant if it exceeds a $\pm 3\%$ threshold. Exceptions are documented in the "Comments" column.

^c The change in % of each variable is synthesised with a single value (-1 for a negative assessment, 0 for an unchanged assessment without marked change and +1 for a positive assessment). These values are then added up and the results determine the general trend for this indicator.

Source: FSO 2012 (report)

Figure 3.7: Communication of assessment results at a grouped level — example United Kingdom

Indicator number and title		Change since 1990 ¹	Change since 2003	Direction in latest year*
1. Greenhouse gas emissions		✓	✓	✓
13. Resource use		✓	✓	✓
18. Waste arisings		⋯	✓	✓
20. Bird populations	Farmland	✗	✗	✓
	Woodland	⊕	✓	✓
	Seabird	⊕	⊕	≈
27. Fish stocks sustainability		✓	✓	✓
28. Ecological impacts of air pollution	Acidity	⋯	⊕	⋯
	Nitrogen	⋯	⊕	⋯
30. River quality	Biological	✓	⊕	≈
	Chemical	✓	✓	✓
32. Economic output		✓	✓	✗
37. Active community participation		⋯	✗	✗
38. Crime		✓ 1991	✓	✓
40. Employment		⊕	⊕	✗
41. Workless households		⋯	⊕	✗
43. Childhood poverty	Before housing cost	✓	⊕	✓
	After housing cost	⊕	⊕	
45. Pensioner poverty	Before housing cost	✓	⊕	✓
	After housing cost	✓	✓	✓

¹ Year as shown if not 1990

Source: DEFRA 2010 (brochure)

Table 3.3: Summary table of the assessment results at a grouped level — all Expert Group countries

Assessment methods communicated at the grouped level	Printed publication or PDF publication available online						Online indicators					
	Belgium	Eurostat	Germany	Netherlands	Switzerland	United Kingdom	Belgium	Eurostat	Germany	Netherlands	Switzerland	United Kingdom
Category 1 assesment methods	V	V	V	V	V	n/a	—	—	n/a	—	V	n/a
Category 2 assesment methods	V	n/a	V	n/a	n/a	n/a	—	n/a	n/a	n/a	n/a	n/a
Category 4 assesment methods	V	V	V	V	V	V	—	—	n/a	V	V	—

V: available

—: not available/unclear

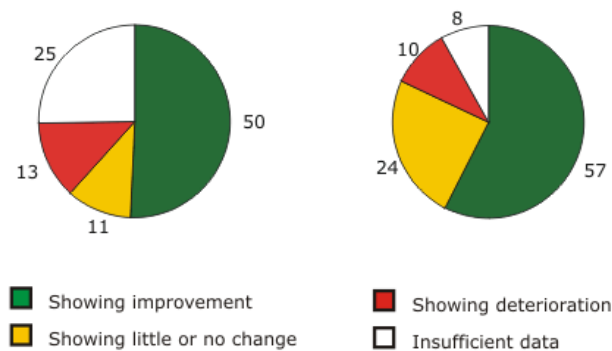
n/a: not applicable

NB: The publications and websites used for this analysis are referenced in Annex 2.

3.3.3.1 Simple, non-interactive aggregated dashboards

This kind of dashboard was introduced at the beginning of the chapter (Box 3.1). A good example of such a dashboard is the pie chart summarizing the assessment results in the brochure 'Measuring progress — Sustainable development indicators 2010' (DEFRA 2010) (see Figure 3.8).

Figure 3.8: Example of non-interactive aggregated dashboard



¹ Based on 99 of 126 measures, comprising 68 indicators

² Or nearest year for which data are available

Source: DEFRA 2010 (brochure)

3.3.3.2 Interactive aggregated dashboards for sustainable development

A prerequisite for developing a sustainable development aggregated dashboard is an underlying **conceptual framework**. This ensures the systematic structure within an aggregated dashboard and represents the tool for assessing the **indicators**, thus providing the building block for the aggregation: the result of the assessment (see Wachtl, Mayerat Demarne, de Montmollin 2010 and Boesch 2011). Aggregated dashboards allow for the aggregation of indicators containing various units by means of the result of their assessment and thereby provide an overall assessment of the situation, while ensuring access to the detail of each indicator.

When developing an aggregated dashboard, achieving transparency is of crucial importance (Wachtl, Mayerat Demarne, de Montmollin 2010). Often a broad variety of indicators with different units are contained in the sustainable development **indicators sets**, reflecting the complexity of the sustainable development concept. By providing an access to each individual indicator (showing underlying data and the methodology employed for indicator assessment), an aggregated dashboard ensures transparency.

Criteria for the communication of assessment results within an interactive aggregated dashboard

An interactive aggregated dashboard should:

- be built upon a solid **conceptual framework** or sustainable development strategy
- allow for a global evaluation of overall assessment of the situation with regard to sustainable development
- provide an overall result of assessment of the situation (by means of symbols)
- give information about the assessment methods used
- provide access to underlying data
- provide the possibility to navigate between the level of individual **indicator** (graph presenting data) and the synthesis level (aggregated level), see Box 3.1.

Switzerland and the Netherlands both developed interactive aggregated SDI dashboards. In the following sections, these interactive aggregated dashboards are presented and reviewed along the lines of these essential elements to illustrate the manner in which SDI interactive aggregated dashboards communicate the **indicator-based assessment results**.

Experiences

Swiss dashboard of sustainable development ⁽²⁴⁾

Homepage of the dashboard ⁽²⁵⁾

The Swiss dashboard is structured according to the 10 key challenges of the Swiss Sustainable Development Strategy. The homepage of the Swiss dashboard (Figure 3.9) presents three primary objectives: 'Social Solidarity', 'Economic Efficiency' and 'Environmental Responsibility'. When the cursor is dragged over the key challenges, the **indicators** of the respective key challenge appear in the corresponding objective. The colour on the left side of the indicator label shows the assessment result of the indicator (green for positive, yellow for neutral and red for negative).

Overview page of the Swiss dashboard

The dashboard presents an overall image of the 10 challenges (Figure 3.10) as well as the possibility of consulting each challenge or **indicator** separately (Figure 3.11). On the red-to-green scales the state of affairs for all 10 key challenges is provided i.e. the synoptic picture of all strategy indicators. Transparency in the dashboard is ensured with the possibility to access the assessment of each of the 10 thematic action areas (challenges). With this possibility the user can ascertain which indicators are used for the **monitoring** of each challenge, and more importantly, how each indicator contributes to the assessment of the key challenges.

⁽²⁴⁾ The description of the Swiss 'Dashboard of Sustainable Development' is based on/paraphrased from Wachtl, Mayerat Demarne, de Montmollin 2010 and Boesch 2011.

⁽²⁵⁾ Available in French and German: <http://www.bfs.admin.ch/bfs/portal/fr/index/themen/21/02/dashboard/01.html>.

Figure 3.9: Swiss dashboard — Homepage

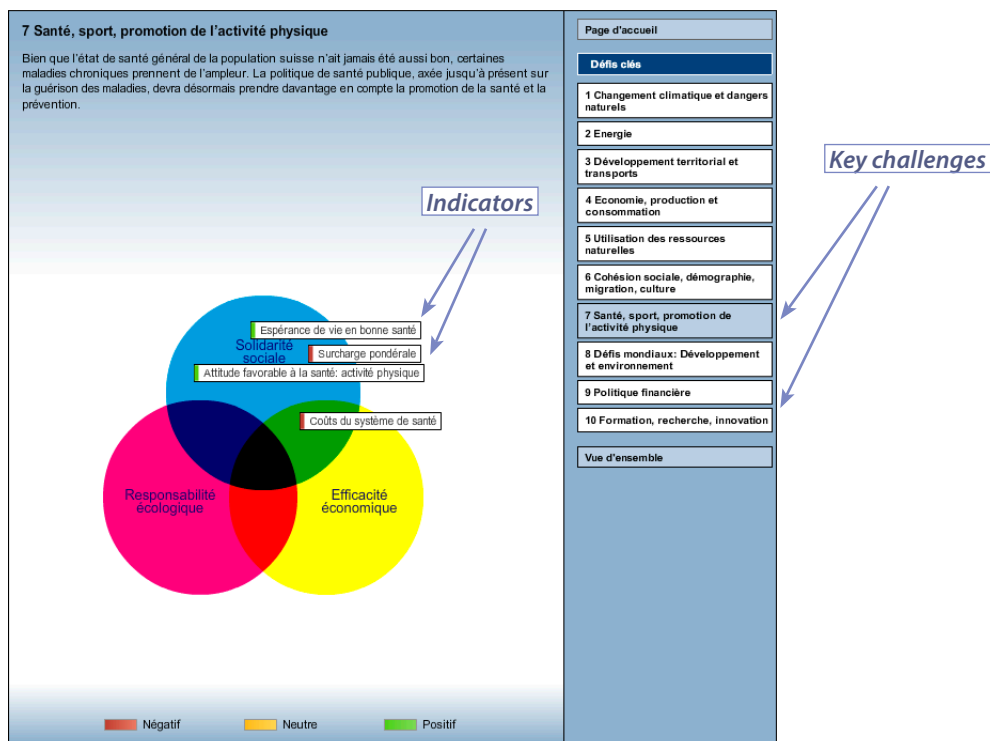


Figure 3.10: Swiss dashboard — Overview page

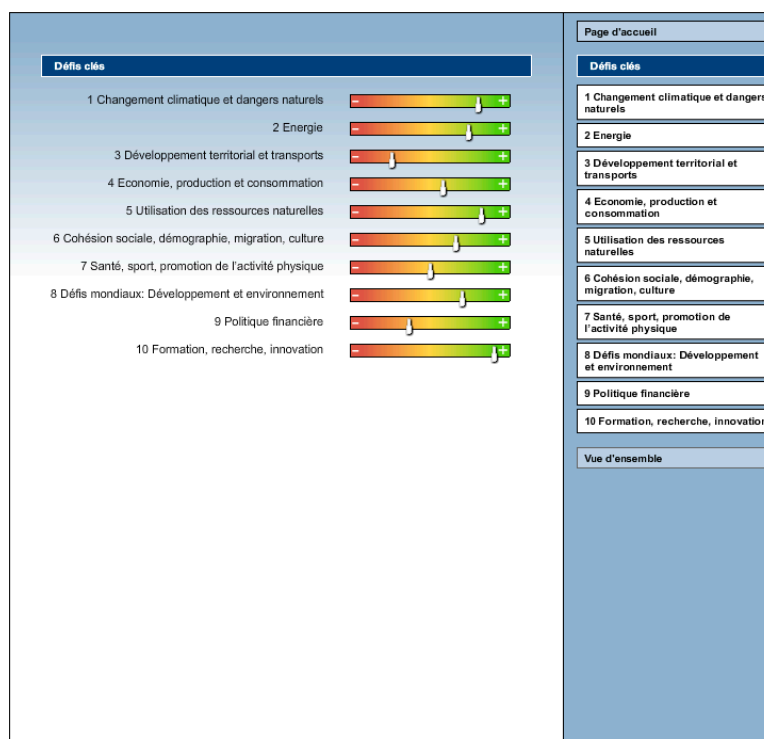
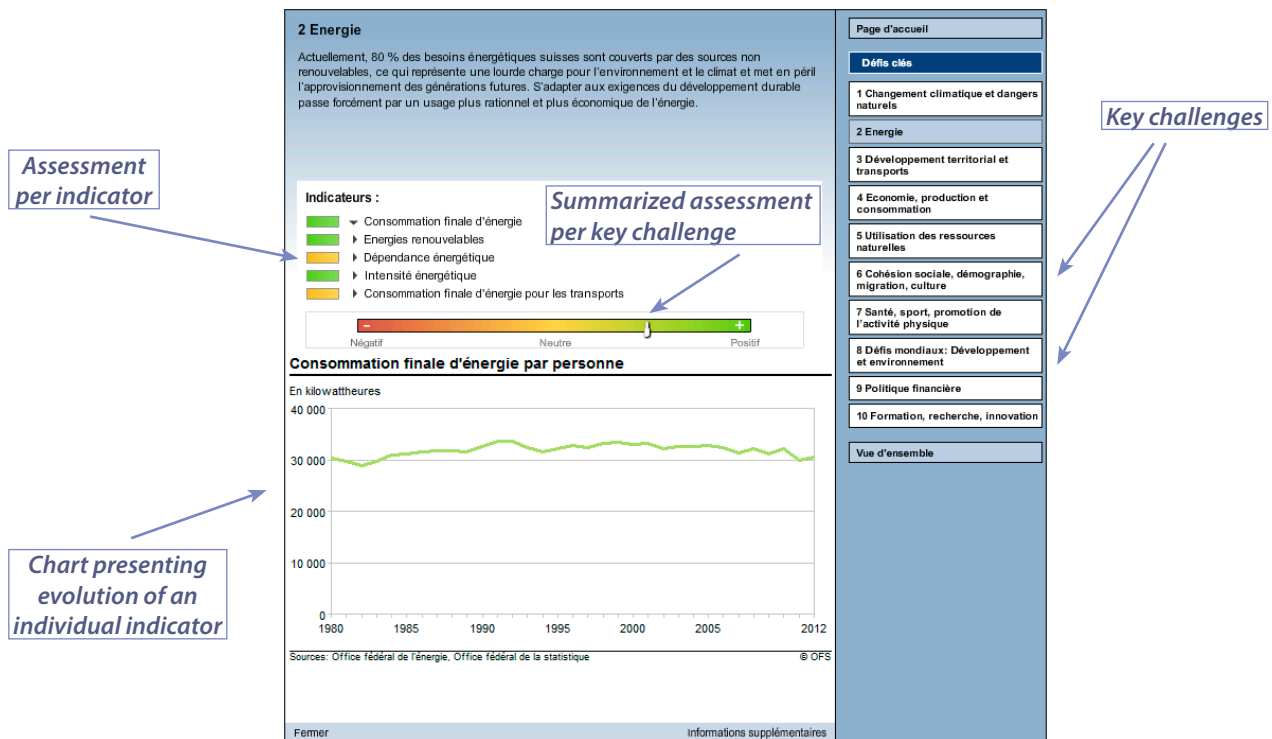


Figure 3.11: Swiss dashboard — Assessment per indicator



Assessment per indicator

The contribution of each indicator is based on a simple mathematical aggregation: if its assessment is positive, it is allocated the value '+1', if it is negative, the value '-1' and if it is neutral, the value '0'. There are five indicators per thematic action area (challenge) and each indicator has the same weight. The overall assessment result per theme is represented by the position of the pointer on a bar which ranges from -5 to +5.

The scale of each key challenge makes use of an animation, whereby the result of the assessment of each underlying indicator is summated one by one. The summation starts in the middle of the scale (0 i.e. neutral) and then moving one unit to the right (+1) in case the first indicator is listed is assessed positively, or one unit to the left (-1) if the assessment is negative. In case the assessment is neutral (0) the pointer doesn't move. The same principle applies to the rest of the indicators. Meta information such as definitions, data sources etc is available on the individual indicator level. Information on the conceptual framework, the assessment methods and the employed symbols is communicated separately on the website.

Dutch dashboard of sustainable development

The Dutch dashboard⁽²⁶⁾ is based on the sustainability framework that follows the guidelines of the Brundtland report as

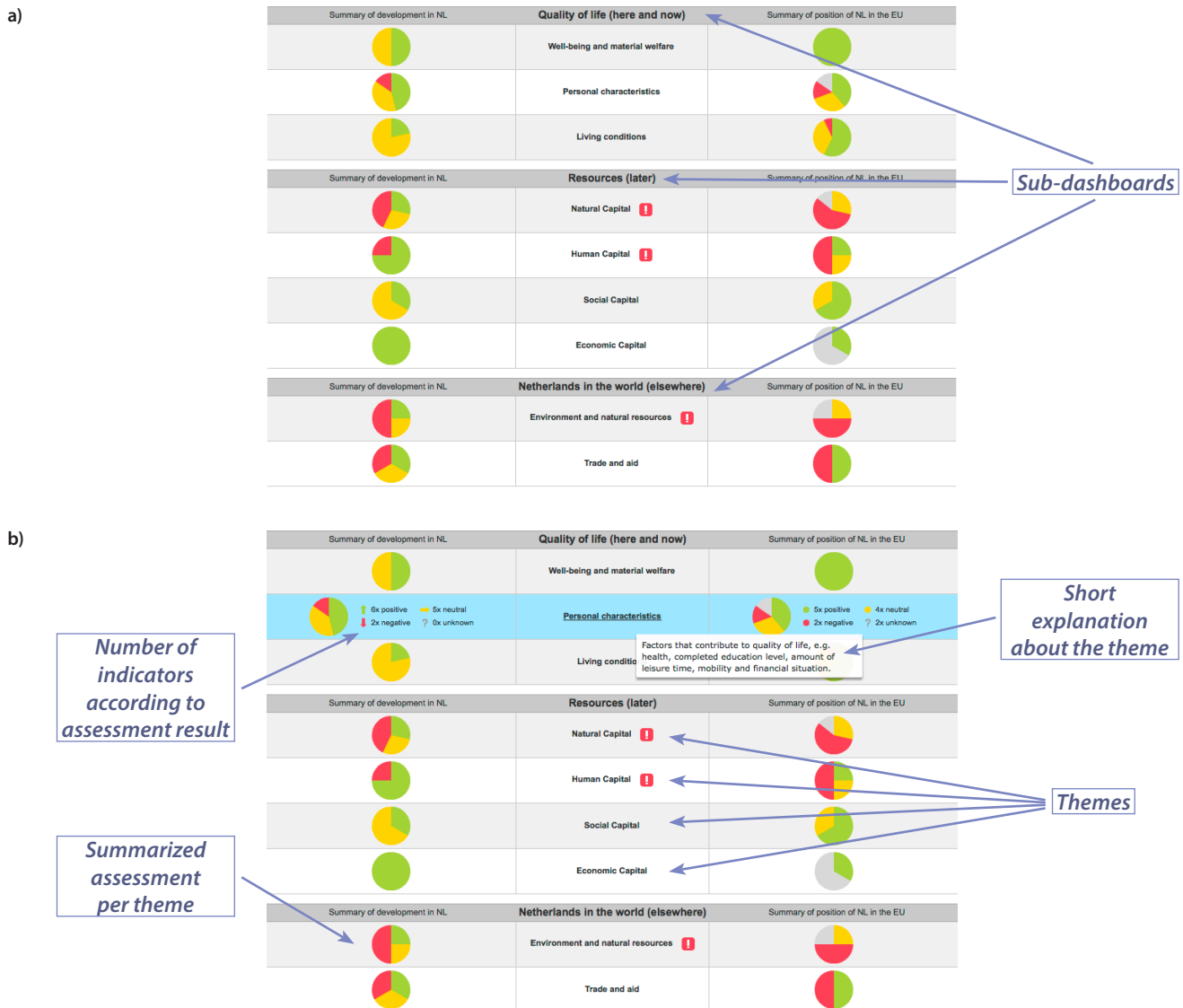
well as the more recent Stiglitz report (Stiglitz, Sen, Fitoussi 2009). It is based on an extended capital approach, which includes present and future aspects of welfare, as well as the international dimension (Smits, Hoekstra 2011). It is therefore subdivided into three sub-dashboards (Figure 3.12a). The sub-dashboard 'Quality of life' illustrates how the Dutch quality of life of the present generation is developing 'here and now'. The sub-dashboard 'Resources' shows whether there is enough capital for future generations to achieve a certain quality of life. And the last sub-dashboard 'Netherlands in the world' illustrates the international dimension of sustainable development and shows the impact of developments in the Netherlands on other countries. The individual sub-dashboards are further split into themes for which indicators have been chosen. The difference between the first column ('Summary of development in NL') and the third column ('Summary of position of NL in the EU') is explained below, under the title 'Assessment per indicator'.

Homepage of the dashboard

Each pie chart shows the distribution of the underlying indicators in a theme regarding their assessment result. Exclamation marks indicate theme areas where developments in the Netherlands give rise to concern. Passing the mouse over a theme brings out more information about it (number of indicators according to their assessment result), and a short explanation about the theme appears in a textbox if the mouse passes exactly over a theme title (Figure 3.12b).

⁽²⁶⁾ Available in English and Dutch: <http://www.cbs.nl/en-GB/menu/themas/dossiers/duurzaamheid/cijfers/extra/duurzame-ontwikkeling.htm>.

Figure 3.12: Dutch dashboard — Homepage



Assessment per indicator

In the first column (Figure 3.13), the developments in the Netherlands since the year 2000 are shown for each individual indicator. The colours (red, green or yellow) are assigned to the symbols (arrows) dependent on the effect an indicator has on sustainability. So, the arrows show whether the observed evolution is increasing or decreasing, and the colours indicate the effect on sustainable development.

In the middle column all indicators belonging to a particular theme are listed. Each indicator has a symbol on the left and the right side that indicate the result of the assessment, either based on the evolution of the indicator (symbol on the left side of the indicator), or on the comparison with other EU countries (symbol on the right side of the indicator).

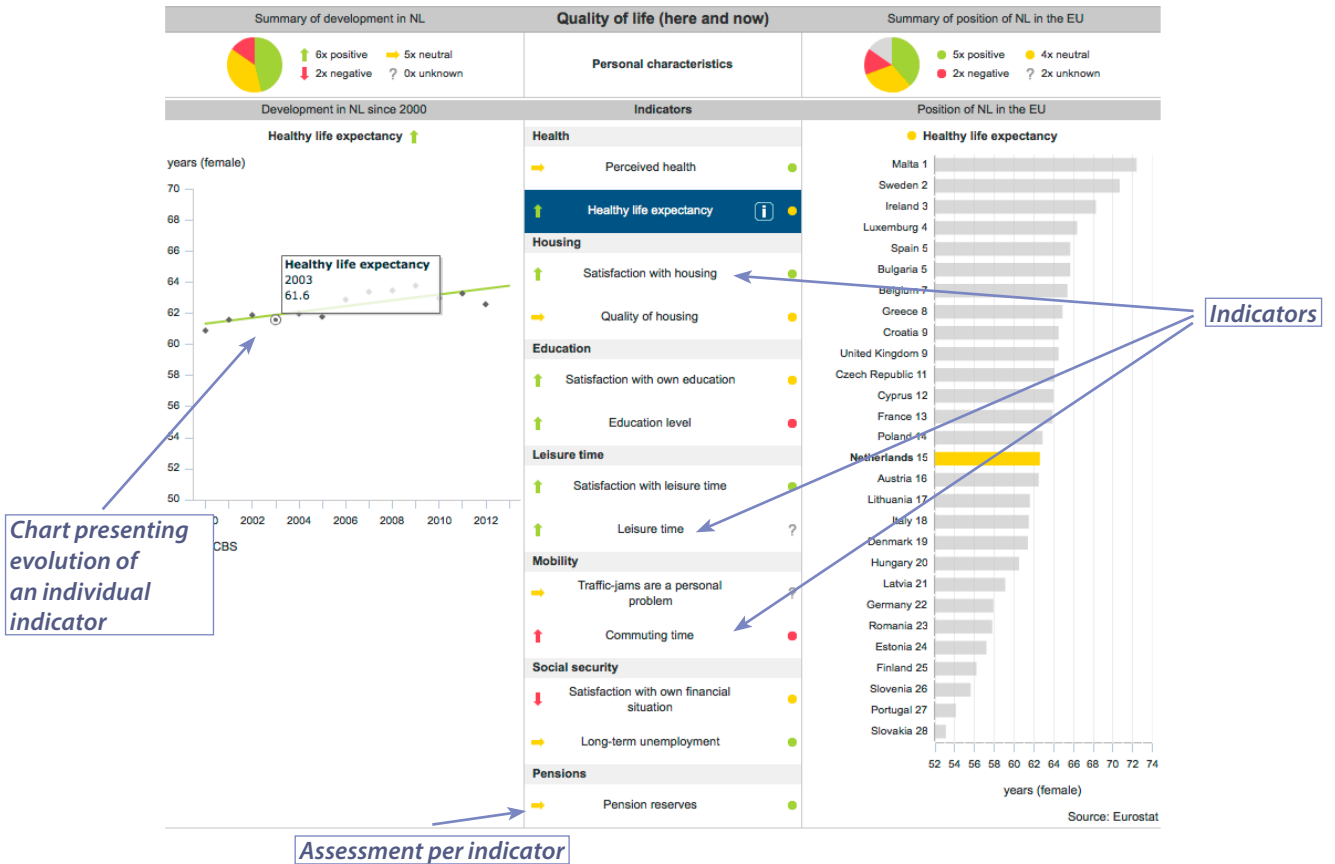
In the third column the developments in the Netherlands compared with other EU countries are presented. This kind of

assessment based on a comparison between countries is not in greater detail since it is not the object of this handbook.

The Dutch dashboard allows simultaneously for the assessment of the development of individual indicators over years and the state (used for international comparison) of an individual indicator from the most recent year where data is available. Besides, this dashboard provides an overview of the overall situation and access to the individual indicators. By showing underlying details, transparency is ensured.

On an individual indicator level, the Dutch dashboard provides metadata information such as definitions, relevance of the included themes and by pointing on a data point the value of an indicator in a particular year. Information on the conceptual framework, the assessment methods and the employed symbols is communicated separately on the website.

Figure 3.13: Dutch dashboard — Assessment per indicator



Discussion of the interactive dashboards of sustainable development

Both the Swiss and Dutch aggregated dashboards provide an overall assessment of the situation by presenting a summary of the assessment results by theme or key challenge. They offer the possibility to navigate between the synthesis level and the individual indicator level. This allows the user to know how each indicator contributes to the overall assessment and gives the user access to detailed information of the indicators (data, meta-information), thus ensuring transparency. There is no direct connection to the conceptual framework behind the dashboard neither to the assessment methodology. This information is presented online, but not directly ‘inside’ the dashboards.

In both the Swiss and Dutch aggregated dashboards, each indicator is attributed an equal weight. This is not based on complex theoretical fundamentals, but rather on a pragmatic approach. As mentioned in Boesch (2011), more sophisticated methods could be used to set up a weighting scheme, for example on the basis of a participatory process, or of scientific publications. This would have an influence on the overall picture given by the aggregated dashboards. However, in both cases, it is clear for the user that the indicators have an equal weight, which ensures transparency.

Boesch (2011) also mentions that while the aggregation is possible thanks to the fact that the indicators are assessed based on their evolution, it is not possible to follow the evolution of the aggregated dashboards. One could imagine that a ‘time function’ could be added to the aggregated dashboards, so that the user could follow the evolution of the assessment of the themes or key challenges. The aggregated dashboards would not be a picture anymore, but a movie showing whether the observed evolutions of the indicators are in line or not in line with their desired evolutions over time. There would be some restrictions, however, due to the fact that there are changes in the indicators that compose the aggregated dashboards (due to lack of data or breaks in time series).

By simplifying the information, the dashboard underlines the weakness associated with the use of indicators (Boesch 2011). For example, users tend to think that the position of the pointer on the bar (Swiss dashboard) means that the current status of the underlying indicators is assessed whereas the assessment is not based on the status of the indicator but on its evolution. This is especially disturbing in the case of indicators for which target values with a target year have been defined and are well known, such as the greenhouse gas emissions indicator. A positive assessment for this indicator does not mean that the Kyoto target has been reached but that the indicator is evolving in the right direction in order to reach the target.

3.4 On the use of symbols





Symbols are used to communicate the results of the indicator-based assessment. The symbols are of many different kinds and differ from traffic lights to weather symbols. Three main systems are used: smileys (stylized facial expressions), colour-coded symbols (e.g. arrows or arithmetic operators in combination with traffic light colours, where green/positive: moving towards sustainability; red/negative: moving away from sustainability and yellow/neutral: no significant changes) and weather icons. Three assessment classes can be distinguished as a starting point: positive, negative and intermediate. The intermediate class can be described as neutral. The application of colours can assist readers in interpreting complex indicators. However, using colour alone (not in combination with a smiley or arithmetic operator) might be a problem for colour-blind persons. Symbols make it much easier to see which indicator is doing well and which not so well, even if it still needs some effort to understand the indicator. Whatever the symbol is, its purpose and description should be clear to understand.

The graphical symbols may have been carefully thought out and tested on a large number of users. However, because of the difficulty of conveying the multifaceted aspects of the evaluation, they still have their limitations. This issue was raised by Hulliger and Lussmann, but has not led to a conclusive result in determining which graphical representation is the most effective for communication (Hulliger, Lussmann 2010).

Traffic light colours printed in black and white may disappear and colour-blind people cannot see the difference. It is important to look carefully for appropriate colour shades (traffic light colours in various tones) that can be distinguished in such circumstances. Attention should be paid to the sufficient disparity in brightness whereby, if colours are printed in black and white, the difference will still be clearly discernible.

In Table 3.4 various symbols applied across the Expert Group countries are presented with respect to their interpretability and common understanding.

Table 3.4: Overview of symbols used for communication of assessment results

Symbol systems	Visual presentation of symbols	Countries
Smileys (in combination with traffic lights colours)		Belgium
Traffic lights colours (in combination with arrows)		The Netherlands
Traffic lights colours (in combination with typographic symbols)		United Kingdom, Switzerland
Weather symbols/icons		Eurostat ⁽²⁷⁾ , Germany

⁽²⁷⁾ Starting from 2013, solid weather symbols are used only if the trend is significant according to the Spearman's rank correlation; if this is not the case, shaded weather symbols are used.

Conclusions and Outlook

4

The Introduction and the first chapter of this handbook describe the context in which **indicator-based assessment** emerged, and the risks associated with synthesizing the information provided by **indicators**. Chapter 1 further details the challenge of going beyond the assessment at the level of individual indicators towards a further synthesis of information. It also looks at how indicator-based assessment fits in policy evaluation, and the role that official statistics play in this respect.

The second chapter suggests a terminology, as well as a classification of the assessment methods that is based mainly on the (non-)availability of predefined **target values**. Indeed, it has traditionally not been the role of organisations responsible for carrying-out indicator-based assessment to define target values and **target years**. They are therefore dependent on this being done by policy-makers. Ideally, however, the definition of target values and target years should be the result of a dialogue between policy-makers and statisticians. The chapter presents the assessment methods that are used by the Expert Group countries in a detailed manner and compares these methods. An important conclusion drawn in Chapter 2 is to avoid relying solely on the result of the assessment (e.g. in the form of pictograms) and to consider the complexity that hides behind by including a broader analysis of the situation in the assessment process.

The third chapter presents first answers to the challenge of going beyond the assessment at the level of individual indicators. This chapter also underlines the importance of communicating in a transparent manner so that the user can understand which elements are taken into account in the assessment process.

There is no straightforward answer to all the issues raised in the handbook. However, this chapter outlines possible paths to follow, and discusses the way forward for indicator-based assessment. Four avenues are presented. The first one addresses how to reinforce the relevance of the indicators and the robustness of their assessment by taking into account the interactions between indicators. The second outlines possible improvements in the communication of the processes and results of indicator-based assessment. The third avenue evokes the necessity to integrate the knowledge on indicator-based assessment in the post-2015 Agenda, especially for the setting-up of Sustainable Development Goals (SDGs). The fourth and last avenue reminds that the experiences described in this handbook can be useful for persons working with indicators outside the field of sustainable development.

4.1 Interactions

The work on SDIs so far was focused on choosing the right **indicators** on the basis of criteria such as quality of data, relevance of the indicators for the issue measured etc ⁽²⁸⁾. A lot of attention has also been paid to the communicability of the

indicators, **indicator-based assessment** being seen as a way to synthesise the information. Further steps have been taken in order to get an overview of the separate assessments for the whole set of indicators, by communicating the assessment results by means of dashboards or aggregated dashboards (see Chapter 3).

Looking at interactions between the indicators of an **indicator system**, or between these indicators and contextual variables, should ideally provide some more information on the reasons of the evolution of the indicators. This should therefore allow a more in-depth analysis of the evolution of the indicators and of the quality and relevance of its assessment. For example, the price of oil, as a contextual variable, is likely to affect the economic indicators one way and the environmental indicators in another. One could also be interested in analysing how smoking or traffic accidents can link to healthy life years.

In the existing indicator systems, the analysis of the indicators often mention interactions that are well established. Similarly, the contextual variables that supplement several indicator systems give extra information that can put the individual indicators in a broader frame (see Chapter 1, section 1.2).

4.1.1 Frameworks and models of interactions are needed

Statistical offices cannot afford to make too many assumptions about the relations between **indicators** or between indicators and other variables. The models they use need to rely on a broad scientific consensus regarding their validity and reliability.

A first step towards enhancing standardisation and fostering research about interactions is to use frameworks or general models in order to structure the **indicator sets** or **systems**. In the field of environmental statistics, the Driving Forces – Pressures – State – Impact – Responses (DPSIR) framework adopted by the European Environment Agency ⁽²⁹⁾ is often used as a structuring element for the environmental indicator systems and for some sustainable development indicator systems. Another example of a general model is the MONET typology used for the Swiss sustainable development indicator system (see FSO 2012, p. 63). A framework can also be an accounting system, such as the UN System of National Accounts (SNA) or the UN System of Environmental-Economic Accounting (SEEA).

Using such frameworks or general models can contribute to the development of the statistical system as a whole: it allows defining in a systematic manner what information is lacking in order to depict the interactions and thus enhance the assessment of the indicators. One way to overcome the lacking information is to collect more data and/or to collect data in a way that allows crosscutting analyses (see section 4.1.2). However, the interactions sometimes first need to be established and proven before they can be measured or illustrated at all. It seems to be particularly a problem in the social dimension and in its

⁽²⁸⁾ According to the frame of reference (see Chapter 2).

⁽²⁹⁾ <http://www.eea.europa.eu/>.

interaction between the environmental and economic dimensions. If complex models of interactions depicting the structure and dynamics of the whole system can be established, indicators can then be derived to reflect important characteristics of the system such as its resilience (see Dahl 2007, p. 170). The system here is understood as covering all the interactions within the society at large — taking into account all kinds of flows and of disparities — as well as the interactions between the society and its ‘substrate’, the environment at large.

4.1.2 Common classifications of data is a prerequisite

Ideally, the statistics that feed into the indicators would be collected so that interactions between economic, social and environmental themes were easy to investigate. This is sometimes possible, but not on a large scale because the classifications used for the collection and/or compilation of official statistics are sector- or theme-specific and are not all based on the same concepts. It is possible to reclassify some of the underlying data so that interactions can be better investigated, but it is necessary to proceed with some caution when using the data for new and more detailed assessments.

The environmental accounts do some of the reclassifications between the economic and environmental statistics. For example, the national uses of resources, environmental pressure, employment, and economic results are all tied together through the classification of economic actors, the government, as well as the households and non-governmental organisations. As different industries, the public sector, civil society and households are identifiable, the system allows for a full coverage of the national actors. It is also possible to use the system for sub-national, sectoral or international analyses, on the premise that the data quality is sufficient for such disaggregation (Eberhardson, Palm, Villner 2007).

Also, in social statistics, there are major efforts to use the same classifications between different statistical areas, in order to improve the possibilities for crosscutting analyses. For example, to make sure that definitions of households are harmonised between different types of social statistics, or that the age groups are possible to align.

4.2 Improving the communication of indicator-based assessment

The communication of the processes and results of indicator-based assessment has so far mainly been carried out by the statisticians in charge of the indicator-based assessment. The communication focuses primarily on the individual indicator level, or at the level of groups of indicators (the so-called ‘dashboards’, see Chapter 3, sections 3.1 and 3.3.2). Three countries

have taken a further step by communicating in an aggregated manner in order to provide an overview of the situation (the Netherlands, Switzerland and the United Kingdom), and two of them (Switzerland and the Netherlands) offer interactive aggregated dashboards. These dashboards allow to navigate between the individual indicator level and the aggregated level, offering the user the possibility to get an overview as well as going deep in the data and meta-data of the indicators.

The communication of the interactions between the indicators and of their influences on the assessment result is worth exploring. A good start could be, for example, to represent in some graphical manner the connections between indicators, and between indicators and contextual variable. In this respect, the collaboration between statisticians and graphic designers may lead to inspiring results, such as the experience made in Switzerland, where graphic designers represented the Swiss SDI system in several manners using the result of the assessment of each individual indicator as building blocks (see Hahn, Zimmermann 2009). The information exchanged at the International Conference on Visualising and Presenting Indicator Systems that took place in March 2005 can also be a good source of inspiration⁽³⁰⁾. The possibility for the user to interact with the indicators, e.g. by choosing its own parameters for carrying out the assessment and comparing the assessment with the ‘official’ one should also be explored in order to take due account of the fact that different individuals or organisations have different views, perspectives and values (see Dahl 2007, p. 171). In this respect, initiatives such as Gapminder⁽³¹⁾ or the OECD Better Life Index⁽³²⁾ could be inspiring, provided one is aware that the scope and goals of such initiatives is different from communicating the process and results of indicator-based assessment.

4.3 Role of indicator-based assessment in the post Rio+20 agenda

As already mentioned, it is not the role of official statistics to define targets, but the one of policy-makers. The dialogue between statisticians and policy-makers needs to be strengthened, in order to ensure that the targets set by policy-makers are actually measurable. This is particularly urgent in the context of the setting of the Sustainable Development Goals (UN 2012, especially paragraphs 250 and 251). The policy-makers would not only benefit from the knowledge of the statisticians concerning what figures are available to measure the targets, but also from their knowledge about the indicator-based assessment methods as well as frames of references and models

⁽³⁰⁾ This conference was organised by the FSO in March 2005 in Neuchâtel, Switzerland. The proceedings are available under <http://www.bfs.admin.ch/bfs/portal/en/index/themen/21/11/visu.html>.

⁽³¹⁾ www.gapminder.org.

⁽³²⁾ www.oecdbetterlifeindex.org.

or typologies. The statisticians could indeed contribute to ensure that all relevant aspects are taken into account and appropriately assessed, e.g. by avoiding focusing only on natural and economic resources, and by also taking into account how efficiently the resources are used, the disparities of access to the resources, and how the different indicators used together interact. They could also contribute to the implementation of the recommendations of the UN System Task Team on the post-2015 UN Development Agenda concerning monitoring and benchmarking, which are based on an analysis of the experiences made with the Millennium Development Goals (MDGs) (UNDP 2012, p. 37–38).

It may not be crucial to define a quantified target together with a target year for all areas covered. In some cases a direction to follow or a growth rate to maintain may be judged as sufficient. In any case, the measurability should be kept in mind at all stages.

4.4 Role of indicator-based assessment outside the field of sustainable development

This handbook is based on the experiences made by practitioners working in the field of sustainable development indicators. Sustainable development being a highly normative field, one has to be very cautious about what assumptions are made and about the transparency. Their awareness of the risk of delivering too simplistic messages and of a misinterpretation of the results combined with the necessity of delivering synthetic messages led these practitioners to take stock of the situation. They hope that the results described in this handbook will also be useful not only for newcomers in the field of indicator-based assessment, but also for practitioners dealing with other types of indicators, such as innovation and education indicators.

Bibliography

- Balthasar 2000:** Balthasar A., *Evaluationssynthesen: Bedingungen, Möglichkeiten und Grenzen*, LeGes – Gesetzgebung & Evaluation, 11(1), 2000.
- Bell, Morse 2003:** Bell S., Morse S., *Measuring Sustainability — Learning by Doing*, Earthscan, London and Sterling, 2003.
- Boesch 2011:** Boesch A., ‘Visual Aggregation of a Sustainable Development Indicators System (MONET)’, *Proc. 58th World Statistics Congress*, Int. Statistical Inst., Dublin, 2011, pp. 3098–3107. (<http://2011.isiproceedings.org/papers/650166.pdf>)
- CBS 2011:** Statistics Netherlands (CBS), *Monitor Duurzaam Nederland 2011*, Den Haag, Heerlen, Bilthoven, 2011 (in Dutch).
- Dahl 2007:** Dahl A. L., *Integrated Assessment and Indicators*, In: Hák T., Moldan B., Dahl A. L., (Ed.): *Sustainability Indicators — A Scientific Assessment*, SCOPE 67, Island Press, Washington, Covelov and London, 2007.
- DEFRA 2010:** Department for Environment, Food and Rural Affairs (DEFRA) and National Statistics, *Measuring progress — Sustainable development indicators 2010*, London, 2010. (http://sd.defra.gov.uk/documents/SDI2010_001.pdf)
- DEFRA 2013:** Department for Environment, Food and Rural Affairs (DEFRA), *Sustainable Development Indicators*, London, 2013. (<https://www.gov.uk/government/publications/sustainable-development-indicators-sdis>)
- DESTATIS 2012:** Federal Statistical Office of Germany (DESTATIS), *Sustainable Development in Germany — Indicator Report 2012*, Wiesbaden, 2012. (<https://www.destatis.de/EN/FactsFigures/Indicators/SustainableDevelopmentIndicators/SustainableDevelopmentIndicators.html>)
- Eberhardson, Palm, Villner 2007:** Eberhardson M., Palm V., Villner M., *Sustainable development indicators based on environmental accounts*, Statistics Sweden, 2007. (http://www.scb.se/statistik/_publikationer/MI1301_2006A01_BR_MIFT0703.pdf)
- European Commission 2004:** European Commission, *Evaluating EU Activities — A practical guide for the Commission service*, Office for Official Publications of the European Communities, Luxembourg, July 2004. (http://ec.europa.eu/agriculture/eval/guide/eval_activities_en.pdf)
- ESSC 2011:** European Statistical System Committee (ESSC), *European Statistics Code of Practice for the National and Community Statistical Authorities*, adopted on 28th September 2011. (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-32-11-955)
- Eurostat 2011:** Eurostat, *Sustainable Development in the European Union — 2011 Monitoring Report of the EU Sustainable Development Strategy*, Luxembourg, Publications Office of the European Union, 2011. (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-31-11-224)
- FPB 2011:** Federal Planning Bureau (FPB), *Twenty Years of Political Commitment to Sustainable Development? — Federal Report on Sustainable Development 2011*, Brussel, 2011. (Summary available in English, full report in French and Dutch) (http://www.plan.be/publications/Publication_det.php?lang=en&TM=41&IS=63&KeyPub=1100)
- Feller-Länzlinger et al. 2010:** Feller-Länzlinger R., Haefeli U., Rieder S., Biebricher M., Weber K., *Messen, werten, steuern. Indikatoren — Entstehung und Nutzung in der Politik*, Zentrum für Technologiefolgen-Abschätzung TA-SWISS, Bern, 2010. (Summary available in English, full report in German) (<https://www.ta-swiss.ch/en/projects/social-and-cultural-ta/indicators/>)
- FSO 2012:** Swiss Federal Statistical Office (FSO), *Sustainable Development Report 2012*, Neuchâtel, 2012. (<http://www.bfs.admin.ch/bfs/portal/en/index/themen/21/22/publ.html?publicationID=4066>)
- Hahn, Zimmermann 2009:** Hahn B., Zimmermann C., *Visualisierung von Indikatorensystemen am Beispiel der Nachhaltigen Entwicklung in der Schweiz*, Forschungsprojekt HKB, Bern, 2011. (<http://www.von-b-und-c.net/designforschung.php?projekt=5&s=en>)
- Hardi, Zdan 1997:** Hardi P., Zdan T. J., *Assessing Sustainable Development: Principles in Practice*, International Institute for Sustainable Development (IISD), Winnipeg, 1997. (<http://www.iisd.org/publications/pub.aspx?pno=279>)
- Hass, Palm 2012:** Haas J., Palm V., *Using the right environmental indicators: Tracking progress, raising awareness and supporting analysis. Nordic perspectives on indicators, statistics and accounting for managing the environment and the pressures from economic activities*, TemaNord, Copenhagen, 2012. (<http://www.norden.org/en/publications/publikationer/2012-535>)
- Hulliger, Lussmann 2010:** Hulliger B., Lussmann Pooda D., *Assessment of Sustainable Development and Environmental Indicators, Technical report School of Business of the University of Applied Sciences Northwestern Switzerland FHNW No 20*, Olten, 2010.

- IISD 2009:** International Institute for Sustainable Development (IISD), *BellagioSTAMP: Sustainability Assessment and Measurement Principles*, Winnipeg, 2009. (<http://www.iisd.org/measure/principles/progress/bellagiostamp/>)
- Kulig, Kolfoort, Hoekstra 2009:** Kulig A., Kolfoort H., Hoekstra R., ‘The case for the hybrid capital approach for the measurement of the welfare and sustainability’, *Ecological Indicators*, Vol. 10, 2010, pp. 118–128. (http://www.cbs.nl/NR/rdonlyres/20D56716-8CFD-4A0F-B372-2AB2085268CE/0/Ecological_indicators_Kuligetal.pdf)
- Parris, Kates 2003:** Parris T. N., Kates R. W., ‘Characterizing and Measuring Sustainable Development’, *Annual Review of Environment and Resources*, Vol. 28, 2003, pp. 559–586.
- Pinter, Hardi, Bartelemus 2005:** Pinter L., Hardi P., Bartelemus P., *Sustainable Development Indicators: Proposals for a Way Forward*, International Institute for Sustainable Development IISD, 2005. (<http://www.iisd.org/publications/pub.aspx?pno=769>)
- Smits, Hoekstra 2011:** Smits J. P., Hoekstra R., *Measuring Sustainable Development and Societal Progress: Overview and Conceptual Approach*, Statistics Netherland, 2011. (<http://www.cbs.nl/NR/rdonlyres/C32647F1-1EBB-4CDF-861C-F80A8BD99CF3/0/measuringsustainabledevelopment.pdf>)
- Stiglitz, Sen, Fitoussi 2009:** Stiglitz J. E., Sen A., Fitoussi J.-P., *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Paris, 2009. (<http://stiglitz-sen-fitoussi.fr/en/index.htm>)
- UN 1992:** United Nations Division for Sustainable Development (UNSD), *Agenda 21*, New York/Geneva, 1992. (<http://sustainabledevelopment.un.org/index.php?page=view&nr=23&type=400&menu=35>)
- UN 2009:** United Nations (UN), *Measuring Sustainable Development*, New York/Geneva, 2009. (<http://www.unece.org/stats/sustainable-development.html>)
- UN 2012:** United Nations (UN), *The Future We Want*, Outcome document of the Rio+20 Summit, 2012. (<http://www.un.org/en/sustainablefuture>)
- UN 2014a:** United Nations (UN), *Conference of European Statisticians recommendations on measuring sustainable development*, New York/Geneva, 2014. (<http://www.unece.org/stats/sustainable-development.html>)
- UN 2014b:** United Nations (UN), *Resolution A/68/L.36 adopted at the 68th United Nations General Assembly on 29 January 2014 — Fundamental Principles of Official Statistics*, New York/Geneva, 2014. (<http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>)
- UN/EC/IMF/OECD/World Bank 2003:** United Nations (UN), European Commission (EC), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), World Bank: *Handbook of National Accounting, Integrated Environmental and Economic Accounting 2003 (SEEA 2003) — Final draft circulated for information prior to official editing*, 2003. (<http://unstats.un.org/unsd/envaccounting/pubs.asp>)
- UNDESA 2007:** United Nations Department of Economic and Social Affairs (UNDESA), *Indicators of Sustainable Development: Guidelines and Methodologies, 3rd Edition*, New York, 2007. (<http://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=107&menu=35>)
- UNDP 2012:** United Nations Development Programme (UNDP), *Realizing the Future We Want for All — Report to the Secretary-General*, New York, June 2012. (<http://www.undp.org/content/undp/en/home/librarypage/poverty-reduction/realizing-the-future-we-want/>)
- Wachter 2010:** Wachter D., *Politischer Nutzen von Evaluationen und Monitoring am Beispiel der Schweizer Nachhaltigkeitsstrategie*, In: Steurer R., Trattning R. (Ed.): *Nachhaltigkeit regieren: Eine Bilanz zu Governance-Prinzipien und -Praktiken*, München, 2010.
- Wachtl, Mayerat Demarne, de Montmollin 2010:** Wachtl J., Mayerat Demarne A.-M., de Montmollin A., *Dashboard of Sustainable Development: Visual aggregation of the Swiss Sustainable Indicators System*, Swiss Federal Statistical Office, Neuchâtel, 2010.

Annexes

Annex 1: Classification of the methods: correspondence table with Hulliger, Lussmann 2010

Section of the Handbook		Method in Hulliger, Lussmann 2010
2.2.1.1.1	Target value and target year available: Linear path to target	Comparison with a course
2.2.2	Target value available, no target year defined: German method 'simple comparison of the last available value (y _t) with the target value (x) taking into account the desired direction.'	Comparison with a target value
2.2.3	No target value defined, rate of change available	Comparison with a heading

Annex 2: Country-specific publications and websites

	Printed publication or PDF publication available online		Online indicators	
	Report ⁽¹⁾	Brochure ⁽²⁾	Dashboard/group of indicators	Interactive aggregated dashboards
Belgium	FPB 2011	—	www.indicators.be	—
Eurostat	Eurostat 2011	—	http://epp.eurostat.ec.europa.eu/portal/page/portal/sdi/indicators	—
Germany	—	DESTATIS 2012	—	—
Netherlands	CBS 2011	—	http://www.cbs.nl/sustainability	http://www.cbs.nl/en-GB/menu/themas/dossiers/duurzaamheid/cijfers/extra/duurzame-ontwikkeling.htm
Switzerland	FSO 2012	—	http://www.bfs.admin.ch/bfs/portal/en/index/themen/21.html	http://www.bfs.admin.ch/bfs/portal/de/index/themen/21/02/dashboard/02.html
United Kingdom	—	DEFRA 2010	http://sd.defra.gov.uk/progress/national/	—

⁽¹⁾ For the countries that publish reports, only these publications were used in the analysis, although several of these countries also publish brochures.

⁽²⁾ Used only for countries that do not publish indicator reports.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:
from the European Union's representations (http://ec.europa.eu/represent_en.htm);
from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
by contacting the Europe Direct service (http://europa.eu/eurodirect/index_en.htm) or
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

(* The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm).



Publications Office



doi:10.2785/36378