

**Improving Mutual Trust amongst European Union Member States in
the areas of Police and Judicial Cooperation in Criminal Matters**

*Lessons from the Operation of Monitoring, Evaluation and Inventory
Mechanisms in the First and Third Pillars of the European Union*

Maastricht University

Faculty of Law

Prof. mr. André Klip

Faculty of Arts and Social Sciences

Dr. Esther Versluis

Josine Polak, MA

Preface

This report is the result of a research project on compliance in relation to findings by European Union monitoring and evaluation mechanisms that function within the scope of EU police cooperation and judicial cooperation in criminal matters. It was commissioned to Maastricht University by the Research Department of the Ministry of Justice of the Netherlands (Wetenschappelijk Onderzoek en Documentatie Centrum).

The report was preceded by a Preliminary Study, by the Directorate of European and International Affairs (DEIA) of the Ministry of Justice, on European Union and Council of Europe monitoring and evaluation mechanisms within the scope of EU judicial cooperation in criminal matters.

Prof. Rick Lawson of Leiden University prepared a report on the mechanisms operating within the Council of Europe. Both projects were presented at a Conference on Monitoring and Evaluation Mechanisms in the Field of EU Judicial Cooperation in Criminal Matters at Maastricht University on 2 and 3 June 2009, which was officially opened by Ernst Hirsch Ballin, Minister of Justice of the Netherlands.

The authors very much benefited from the guidance of the supervisory committee chaired by Prof. Olivier De Schutter (Université de Louvain). The committee also consisted of Gisèle Vernimmen- van Tiggelen (formerly EU Commission), dr. Marlèn Dane (Ministry of Justice), Adriëne Boerwinkel (Ministry of Justice), John Morijn, later replaced by Jasper Krommendijk (Ministry of Justice) and Corine van Ginkel (Research Department of the Ministry of Justice).

We are indebted to Steven Freeland (University of Western Sydney), who undertook the task of correcting our English. Our work was further supported by the efforts of our student assistant, Charlotte Hermans.

This report was completed on 31 August 2009.

André Klip
Esther Versluis
Josine Polak

Contents

Preface	p.1
Contents	p.3
1. Introduction	p.5
1.1 Background of the study	p.5
1.2 Further elaboration of the research topic	p.6
1.3 Research questions and outline of the report	p.8
2. Setting the scene: on compliance in the European Union	p.11
2.1 What is compliance and why is it relevant?	p.11
2.2 Actors involved in compliance in the EU's first pillar	p.14
2.3 Difficulties related to EU level oversight functions	p.16
2.4 Compliance with legislation with evaluation and monitoring mechanisms	p.19
2.5 Analyzing compliance problems in the first pillar	p.20
2.6 Analyzing compliance problems in the third pillar	p.24
2.7 Lessons learned	p.33
3. Setting the scene: monitoring and evaluation mechanisms	p.35
3.1 Learning from the peer review mechanism used in the OECD	p.35
3.2 Monitoring and evaluation mechanisms in the first pillar of the EU	p.41
3.3 Existing evaluation mechanisms in the third pillar on criminal law	p.47
4. Monitoring and evaluation mechanisms at work in the first pillar	p.53
4.1 Characterizing the European Employment Strategy	p.54
4.2 From Peer Review Programme to Mutual Learning Programme: how did we get there?	p.56
4.3 The Mutual Learning Programme: core elements	p.57
4.4 Assessing the effectiveness of the MLP	p.59
4.5 Lessons learned: how to improve the effectiveness of the MLP?	p.67
4.6 Lessons confirmed – Part I: evaluation mechanisms in the anti-discrimination field	p.68
4.7 Lessons confirmed – Part II: the Stability and Growth Pact and the 'Brussels I Regulation'	p.76
5. Monitoring and evaluation mechanisms at work in the third pillar	p.79
5.1 Analysis of the existing evaluation mechanisms	p.79
5.2 Methodology of evaluation: more room for social scientists	p.93
5.3 Lessons learned	p.94
6. Synthesis and recommendations	p.103
6.1 Lessons learned I – From the more abstract context...	p.103
6.2 Lessons learned II – ...to the more pragmatic functioning of evaluation and monitoring mechanisms – possible scenarios	p.105

6.3 To conclude: achieving uniformity while respecting diversity	p.106
Annex – Overview of compliance factors	p.109
7. Executive summary	p.117
8. Leesvervangende samenvatting	p.121

1. Introduction

1.1 Background of the study: why study the monitoring and evaluation mechanisms?

Compliance does matter, although this is not always self-evident.¹ Whereas this truism by necessity applies generally for all three pillars of the European Union (EU or Union), the focus of the present research will be on the issue of compliance with the monitoring and evaluation mechanisms employed within the first and third pillars. The ultimate aim of this research is to arrive at practical recommendations on how to improve compliance with these mechanisms within the third pillar.²

The problem underpinning this research project is the fact that, within the third pillar, the degree of compliance with legal instruments is unsatisfactory. This endangers the rule of law in two respects: it frustrates the efforts and needs of law enforcement and the fair administration of justice (both in a national setting as well as in relation to international cooperation); and, given the divergent status of implementation in the Member States, it seriously distorts the protection of the rights of citizens. However, law enforcement and citizens' rights are recognized as fundamental elements in the Treaty on the European Union: Article 4 TEU states that the EU will provide its citizens within the 'area of freedom, security and justice' with a high level of security – something that naturally implies that both elements are crucially important for making the internal market work. Article 6 TEU further emphasizes the essential character of protecting the fundamental rights of citizens, as embodied in the European Convention on Human Rights (ECHR) in the context of the Union.³

If better compliance with these quintessential principles is to be achieved, the formulation of a common criminal policy – which has yet to be concluded, despite the prominent attention given to combating crime – could be helpful, as it could focus the attention on the grand area of the rule of law *per se*, of which the individual legal instruments that are in place at the current moment are a part.

Finding ways to enhance compliance appears to be very necessary indeed at the current moment, since cases of non-compliance seem to abound. Framework Decision 2002/584 on the European Arrest Warrant constitutes a clear exemplary case of widespread non-compliance. The follow up process regarding this legislative act, which constitutes the

¹ In the words of Roth: L'évaluation est essentiellement une démarche opérationnelle, et elle doit amener à des résultats." R. Roth, Politiques pénales et évaluation: des défis renouvelés, in: A. Weyembergh and V. Santamaria, The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings, Bruxelles 2009, p. 47.

² The initial notice of the Ministry of Justice also speaks of inventory mechanisms. As such, mechanisms mainly refer to the process of 'taking stock' – and relate less to actually taking action to improve situations of non-compliance, compared to evaluation and monitoring mechanisms. We have therefore decided to concentrate on evaluation and monitoring mechanisms. The concept of 'inventory mechanisms' is not frequently used within the EU context.

³ Article 6, paragraph 2 TEU (after entering into force of the Treaty of Lisbon) provides that the Union shall accede to the ECHR.

first Framework Decision in criminal law based on the principle of mutual recognition, demonstrates quite some reluctance among the Member States to cooperate within the area of criminal matters and, as such, a widespread refusal to comply with (some of the) provisions laid down in the Decision. It is very clear that for meaningful cooperation between the Member States to be achieved, the problem of non-compliance must be addressed.

Fortunately, it is not necessary to wait for the Treaty of Lisbon to enter into force before taking action and, in fact, the Stockholm Programme that was established this year may be a perfect opportunity to address problems of non-compliance in the area of third pillar cooperation. The support from the Member States for an initiative by the Netherlands Ministry of Justice to strengthen mutual trust within the domain of police and judicial cooperation in criminal matters, as expressed at an informal meeting of the Ministers of Justice early this year, is very encouraging in this respect,⁴ as mutual trust must be considered to be the necessary foundation on which cooperation can take place.

In order to be able to strengthen mutual trust within the areas mentioned, it is imperative to examine – and subsequently strengthen – the functioning of monitoring and evaluation mechanisms currently existing within the third pillar of the EU. If this type of mechanism functions effectively, the grounds for Member States to distrust others will recede. In other words, improving such mechanisms helps to enhance the degree of mutual trust amongst Member States, as they will be more likely to actually make Member States deliver on their commitments.

In order to produce recommendations on how to actually improve the respective mechanisms, a careful examination of the content and process-related factors that might hinder or stimulate compliance with third pillar mechanisms is required. Such an examination necessitates a close investigation of the actual organization (approach and working method chosen, and constellation of actors involved) of these monitoring and evaluation mechanisms, and the way in which this particular organization influences compliance by Member States.

It is self-evident that any plans intended to improve actual compliance with monitoring and evaluation mechanisms within the domain of criminal matters may benefit from an examination of the experiences of such mechanisms in the first pillar, where significant and longstanding resort to such mechanisms is to be found. The recommendations to be delivered by the researchers will, therefore, be based on a study on the functioning of monitoring and evaluation mechanisms within both the first and the third pillar.

1.2 Further exploration of the research problem

An important point to note is that part of the concerns relating to the unsatisfactory degrees of compliance relate to doubts as to the degree to which Member States have

⁴ Report on the Informal Meeting of the Ministers for Justice and Interior on 15 and 16 January 2009.

implemented EU third pillar legislation, as well as principles related to the rule of law.⁵ Although over the last 15 years the Commission, the Council and the Member States have developed mechanisms of monitoring and evaluation to improve compliance within the third pillar, it may be that the absence of a binding monitoring system and the accompanying infringement procedures that exist within the first pillar also influences the other monitoring mechanisms. One initial suggestion would be to introduce the possibility of initiating infringement proceedings for third pillar issues,⁶ but it is important to first consider and learn from the problems and difficulties of using these proceedings in the first pillar.

If we move to the existing monitoring and evaluation mechanisms, a first glance would suggest that these are predominantly of an administrative nature. Their general feature – and this is from where (part of) the problem of non-compliance derives – is that they are fragmentary, incoherent and inconsistent.⁷ In other words, the existing mechanisms appear insufficient and inefficient, in the sense that they do not guarantee implementation of the common binding norms of the third pillar. The resulting situation of uneven compliance throughout the EU is further exacerbated by the fact that, due to the idea of ‘variable geometry’, not all third pillar legislation is actually applicable to all Member States, and that limiting declarations are made by Member States to third pillar legislation.⁸

How do States cooperate within this overall context of uneven compliance? How do States arrive at truly meaningful cooperation if, consequently, any mutual trust they may have in another Member State’s legal system, or in its implementation, is incomplete?⁹ As the principle of mutual recognition is partially founded on the idea that the Member States share common ground regarding the fundamental principles of their legal systems, we should approach these questions by first identifying whether that assumption can actually be considered to hold true. Indeed, it is relevant to identify which foundational principles – the rule of law, the protection of human rights, the principle of legality, the principle of ‘subsidiarity’, the principle of proportionality and other principles common to all criminal justice systems – are shared by Member States.

The result of such an analysis should clarify whether the assumptions regarding the idea that Member States indeed share common ground are justified and, if not, why they are not. The importance of carrying out this analysis relates to the fact that recommendations on how to improve the effectiveness of monitoring and evaluation mechanisms can only

⁵ See G. Vernimmen-Van Tiggelen and L. Surano, *Analysis of the future of mutual recognition in criminal matters in the European Union*, Université Libre de Bruxelles 2008.

⁶ Although this is provided in the Treaty of Lisbon, one must note that, for the old third pillar acquis, it will take another five years after entering into force of that Treaty for them to finally fall under the umbrella of the Court and Commission.

⁷ André Klip, *European Criminal Law, An Integrative Approach*, Intersentia Antwerpen 2009.

⁸ This is a situation that will worsen when the Treaty of Lisbon enters into force.

⁹ Of all Framework Decisions on mutual recognition, for example, only the Framework Decision 2002/584 on the European Arrest Warrant has been implemented in all 27 Member States.

be made if the standard against which the existing mechanisms to assess effectiveness must be measured is made transparent.¹⁰

Although such questions are of particular relevance in a context where cooperation in criminal matters amongst the Member States is expected or required, they are also relevant in the other two broad areas of the third pillar framework, namely substantive criminal law and criminal procedure. This is so even if one has to keep in mind that, although most of the legislation in the third pillar relates to cooperation, there is also an impressive *acquis* on substantive criminal law, as well as a few instruments that deal with criminal procedure. Again, questions of mutual trust and mutual recognition figure as more or less important in all of these areas, something that makes indispensable a proper analysis focused on these issues.

Returning to the particular research question that is the subject of this report, it should be emphasised that its very relevance derives from the idea that, while evaluation and monitoring mechanisms may greatly contribute to enhancing compliance amongst the Member States, the third pillar currently clearly lacks an effective monitoring system. When ratified, the Treaty on the Functioning of the European Union will provide a formal framework for such an ‘implementation scheme’ (Article 70 TFEU).¹¹ However, given the uncertainty as to the moment of ratification of the Treaty, the five year transition period after the entering into force of the Treaty of Lisbon regarding third pillar legislation, as well as the increasing number of situations in which legislation is not applicable to all Member States,¹² it is wise to already begin to address the current situation by improving the functioning of the existing evaluation and monitoring system.

1.3 Research questions and outline of the report

Based on the above reflections, we propose the following research questions:

1. What content and process related factors hinder (better) compliance with monitoring and evaluation mechanisms within both the first and third pillars of the EU?
2. To what extent does the degree of compliance depend on the approach and working methods used, and the constellation of actors involved in these monitoring and evaluation mechanisms?
3. What adjustments to monitoring, evaluation and inventory mechanisms might improve the follow up to the outcomes of such mechanisms by the Member States?

Based on these questions, the report consists of the following structure:

¹⁰ André Klip, *European Criminal Law, An Integrative Approach*, Intersentia Antwerpen 2009, Chapter 4.

¹¹ See Stine Andersen, *Sovereignty and the emergence of non-binding peer review within the EU*, on file with the authors.

¹² This is referred to as enhanced cooperation, currently provided for in Articles 40-40B TEU.

1. We have made an assumption that, in order to arrive at valuable recommendations on how to improve compliance in the third pillar, factors beyond those merely related to the monitoring and evaluation mechanisms *per se* have to be taken into account. We will begin by providing a broad overview both of relevant contextual factors, at both the EU and national level, that influence compliance in general, as well as more specific mechanisms that stimulate compliance in the first pillar in general (chapter 2);
2. We will analyze existing monitoring, evaluation and inventory mechanisms within the first pillar, thereby focusing on those employed in the domain of employment policy. The results of this case study will be supported by additional cases that are exemplary for how evaluations are undertaken in the first pillar (chapters 3 and 4);
3. We will analyze existing monitoring, evaluation and inventory mechanisms in the third pillar (chapters 3 and 5); and
4. Based on a combination of both analyses, we will provide a synthesis of our findings and produce recommendations on how to improve compliance with monitoring and evaluation mechanisms in the third pillar (chapter

2. Setting the scene: on compliance in the European Union

For the purpose of improving compliance with third pillar mechanisms, a study of the functioning and effectiveness of monitoring and evaluation mechanisms employed within the first pillar of the EU, where great use has been made of such mechanisms, is a logical first step. Indeed, based on experience with evaluation and monitoring mechanisms in other international organizations (in particular within the OECD, WTO and IMF), an interest in the use of such ‘learning mechanisms’ has grown steadily in the EU. The introduction of such mechanisms is most developed and visible in the first pillar, primarily (although not exclusively) in the domains of economic and employment policies, where the principle of governance through ‘policy coordination’ entered the scene during the 1990s. This principle ‘refers to the process through which Member States agree to meet common European concerns and objectives whilst fully preserving their competences to legislate in the respective policy areas’,¹³ and was thus intended to accelerate the process of European integration, without encroaching upon the sovereignty of EU Member States. It is self-evident that any plans to improve actual compliance with the monitoring mechanisms within the domain of criminal matters may benefit from an examination of the experiences with such mechanisms in the first pillar.

2.1 What is compliance and why is it relevant?

Before turning to an overview of the main factors that stimulate and/or hinder compliance, however, it is relevant to come to grips with the term ‘compliance’ in general. Compliance refers to the extent to which ‘agents act in accordance with and fulfilment of the prescriptions contained in (...) rules and norms’.¹⁴ In the example of the EU, compliance thus refers to the extent to which the Member States act in accordance with the provisions of the Treaties, secondary legislation (regulatory measures such as regulations, directives and decisions) and/or soft-law instruments. Compliance is a relative measure. The presence of compliance is hard to prove. Compliance problems seem to occur particularly when the goals are high and demanding. Regulations that require little of States will be relatively easy to comply with.

Many empirical studies on compliance with international law are found in international regimes other than the EU - for example, the WTO, IMF, and GATT. These studies clearly show a correlation between the demands and ambitions of international regimes and their compliance rates; the more demanding a regime, the less likely that its rules will be complied with and, conversely, the less demanding a regime, the more likely that its

¹³ Hartwig, I., and C. O. Meyer, *Towards deliberative network governance? Theorising socio-economic policy co-ordination in the European Union*, 2002. Available at <http://www.govector.org>.

¹⁴ Checkel, J. T., ‘Why Comply? Constructivism, Social Norms and the Study of International Institutions’. ARENA Working Papers, 99/24, Oslo: Advanced Research on the Europeanisation of the Nation-State, 1999.

rules will be complied with.¹⁵ Compliance rates in the human rights area seem to particularly demonstrate this correlation between ambitious goals and the lack of results.¹⁶

Compliance does not, however, equal effectiveness. It is argued that low compliance with demanding rules does not say anything about whether or not this is any more or less effective than high compliance with less demanding rules. In this report, we use the following definition of **compliance**:

the reference is to whether Member States comply with the legal standard that is set, whether the standard is defined in an international treaty, in secondary legislation, or in other, soft-law instruments; generally compliance will require both formal and practical implementation:

where mention is made of **formal implementation**, the reference is to the adoption of legal or regulatory instruments that adapt the regulatory framework to the requirements set by international standards (for example, copying a directive or a framework decision into domestic law); and

where mention is made of **practical implementation**, the reference is to the effective enforcement of existing legislation, in order to ensure that it influences the behaviour of the persons regulated (see also Table 1 below).

In order to establish whether compliance takes place in practice, and thus to ensure that all Member States actually adhere to the international agreements, the EU has developed various **evaluation and monitoring mechanisms**. Establishing a clear-cut differentiation between both sorts of mechanisms is difficult. Hogwood and Gunn¹⁷ differentiate between the two in the following manner:

Evaluation refers to the process of ‘determining whether a policy has been successful in achieving desired outcomes’. ‘Monitoring requires collecting information about the extent to which program goals are being met’, but ‘monitoring is not just about information collection. It requires decisions about what *action* will be taken if performance unduly deviates from what is desired.’

¹⁵ ‘No state ever failed to comply with either the Convention Concerning the Protection of the World Cultural and National Heritage or the Convention on Wetlands of International Importance since they require nothing more than the optional registration of cultural sites and wetlands at the discretion of the state. Nor have the vast majority of the thirty-nine states that are signatories to the International Convention for the Regulation of Whaling, but which have never done any whaling, failed to comply with that agreement’). Downs, G. W., and A. W. Trento, *Conceptual Issues Surrounding the Compliance Gap*, International Law and Organization, Ed. E. C. Luck and M. W. Doyle, New York, Rowman & Littlefield, 2004, p. 19 - 40.

¹⁶ Downs, G. W., and A. W. Trento, *Conceptual Issues Surrounding the Compliance Gap*, International Law and Organization, Ed. E. C. Luck and M. W. Doyle, New York, Rowman & Littlefield, 2004, p. 19-40.

¹⁷ Hogwood, B. W., and L. A. Gunn, *Policy Analysis for the Real World*, Oxford, Oxford University Press, 1984.

It thus becomes clear that both evaluation and monitoring go beyond mere ‘stock taking’. It is clear that it also involves a process of attaching values or opinions to the information that is collected. The above definitions seem to indicate that monitoring goes one step further as compared to evaluation, given the express reference to action taken to undo unwanted behaviour.

Correspondingly, the degree of **effectiveness** of these mechanisms is broadly understood to refer to

whether the system considered brings about the desired effects, both in terms of compliance and formal and practical implementation.

More specifically, an important element in these evaluation and monitoring mechanisms is that Member States *learn* from each other’s experiences. These mechanisms, in other words, aim to stimulate compliance via **policy learning**. The cognitive dimension of learning is very much linked to the practical effect that is expected, namely actual policy transfer/change and thus compliance. A definition of policy learning would include the following elements:

Policy learning is ‘a deliberate attempt to adjust the goals or techniques of policy in response to past experience and new information. Learning is indicated when policy changes as a result of such a process’.¹⁸

Generally, in addition to the effectiveness of evaluation mechanisms, the **efficiency** of such mechanisms is also considered to be an important element of analysis. Efficiency concerns the broad question of

whether the system considered meets a cost-benefit analysis; in other words, whether its benefits outweigh its costs, and whether there would exist evaluation systems that, with less investment in human or financial resources, would achieve the same or even better results.

A last important notion is the concept ‘**rule of law**’, which will be defined as follows in this report:

‘The rule of law is a principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards.’¹⁹

¹⁸ Hall, P., Policy Paradigms, Social Learning and the State, The Case of Economic Policymaking in Britain, Comparative Politics, 1993, p. 278.

¹⁹ Guidance Note of the Secretary-General, UN Approach to Rule of Law Assistance, 14 April 2008.

2.2 Actors involved in compliance in the EU's first pillar

In recent years, particularly since the European Commission published its White Paper on European Governance in 2001, attention on compliance with EU policy has been increasing. As stated in the White Paper, compliance is a necessity 'not only for the sake of efficiency of the internal market but also to strengthen the credibility of the Union'.²⁰ The Commission thus makes it clear why compliance with EU policy is a relevant topic. At the same time, it seems to suggest that there are reasons to believe that problems may occur at this stage.

This suggestion is further highlighted in a communication titled 'A Europe of results', that stresses that 'it is necessary to attach high priority to the application of law, to identify why difficulties in implementation and enforcement may have arisen and to assess whether the present approach to handling issues of application and enforcement can be improved'.²¹ A political organization that produces legislation that is not complied with has a problem with effectiveness, which in the long run has a direct impact on its legitimacy and credibility. As Williams states: 'the EU is liable to appear as a travesty of governance, regulation without implementation'.²²

Although relevant in all political systems, compliance is particularly important in the EU. While EU policy is centrally decided upon in Brussels, it is to be carried out at the Member State level. This potentially leads to 27 different 'practices'. Whenever Member States fail to implement EU policy, the European Commission can intervene, in order to try to ensure that they begin to comply; a situation that is unique for an international organization.

²⁰ Commission of the European Communities (CEC), European Governance: A White Paper, COM (2001) 428 final, Brussels, 25 July 2001, p. 5.

²¹ Commission of the European Communities (CEC), A Europe of Results - Applying Community Law, COM (2007) 502, Brussels, 2 September 2007, p. 2.

²² Williams, G., *Monomaniacs or Schizophrenics?: Responsible Governance and the EU's Independent Agencies*, Political Studies, 2005, p. 88.

*Table 1: Notions of implementation in the European context*²³

Member State level	European level
✓ Formal implementation, i.e. transposition: the ‘law in the books’	✓ The Commission (in cooperation with the ECJ) as the ‘guardian of the Treaties’
✓ Practical implementation, i.e. enforcement by regulators and compliance by the regulated: the ‘law in action’	✓ European agencies
	✓ Direct implementation by the Commission ²⁴
	✓ Comitology ²⁵

In the context of the EU, different people potentially refer to different aspects when talking about ‘compliance’ in the first pillar. Examining this topic thus initially requires a very clear reference to those actions that are being discussed and by which actors. On the European level, the European Commission and the European Court of Justice (ECJ) in particular have responsibilities related to implementation. At the Member State level, a variety of actors come into play; the most prominent being Ministries, inspectorates and ‘the regulated’.

Secondly, it is important to be aware that, while monitoring and evaluation mechanisms may be the prime tool to stimulate and improve compliance in the third pillar, the situation in the first pillar is different. In particular, a more extensive use of binding legislation (in the form of regulations, directives and decisions) creates a different environment, in which compliance is to take place. This particular environment features a more active role for actors (the Commission and ECJ) that currently play a less dominant role in the third pillar. However, greater knowledge of the workings and functioning of these actors can nevertheless provide insight into how and in what direction monitoring and evaluation mechanisms may be improved.

According to Article 10 TEC, it is the Member States who need to ensure fulfilment of European commitments. This does not imply, however, that they are free to ‘do as they please’. According to Article 211 TEC, both the Commission, as ‘the guardian of the Treaties’ and the ECJ are responsible for ensuring that ‘the provisions of th[e] Treaty and the measures taken by the institutions pursuant thereto are applied’. When confronted with non-implementation – for example, incorrect transposition or a lack of application – the Commission can institute infringement procedures (Articles 226 and 228 TEC) and

²³ Versluis, E., Even rules, uneven practices: Opening the ‘Black Box’ of EU law in action. *West European Politics*, 2007, 30(1), p. 50 - 67.

²⁴ In certain policy fields, such as, for example, the administration of the EU’s humanitarian aid programme, or the common fisheries policy, the Commission has the task of directly overseeing the application of European legislation. Most important is the field of competition (Article 85, TEC), where the Commission can, for example, directly prohibit mergers from taking place. These exceptions were put in place in fields where differing interests between Member States are likely to arise, and where, arguably, we need the Commission as an ‘objective’ partner.

²⁵ Article 202 TEC refers to a form of implementing powers of the Commission. This article refers to activities commonly known as ‘comitology’, during which committees composed of Member State officials advise, block or approve the Commission’s proposals for implementing – filling in the details of, and incomplete gaps in – EU legislation.

issue a ‘letter of formal notice’, then a ‘reasoned opinion’, followed by a referral to the ECJ. In the last instance, financial sanctions can be imposed.

Over the years, a growing number of infringement proceedings have been initiated: in recent years, an average of about 1,700 letters of formal notice have been issued, leading to an average of 550 reasoned opinions and 200 referrals to the ECJ per year.²⁶ Given that around 70% of the complaints regarding non-implementation that reach the Commission are actually already solved before a letter of formal notice is sent, and that of all cases that end up at the ECJ, the Commission wins about 90%, we could conclude that the Commission is relatively successful as a guardian or ‘watchdog’ of the Treaties.

This is also reflected in the high percentages of directives that are transposed.²⁷ For years, more than 99% of all directives are transposed. Over the last couple of years, some of the new Member States lead the list in relation to transposition; Latvia and Slovakia particularly seem to be diligent in this regard.

A relative new ‘compliance-related activity’ at the European level refers to European agencies. Since the 1990s, we can observe an increasing trend in delegating specific tasks to independent agencies within the EU. By agencies we mean a ‘variety of organizations – commissions, boards, authorities, services, offices, inspectorates – that perform functions of a governmental nature, and which often exist outside the normal departmental framework of government’.²⁸ Currently, there are 24 agencies in the first pillar, and three in both the second and third pillars, a total of 30 in all. Most can be classed as ‘regulatory’ agencies and some within this category have tasks that relate to implementation. Agencies in the field of transport are particularly involved in implementation. For example, the European Maritime Safety Agency (Lisbon) and the European Aviation Safety Agency (Cologne) – in order to ensure safety in their respective fields – both have the possibility to conduct inspections in Member States.²⁹

Currently, the number of agencies with such far-reaching tasks is limited, and the tasks remain restricted to ‘inspecting the inspectors’ (that is, inspecting the work of the national inspectorates but not the actors actually being regulated). The future will tell whether this will become a new trend in the EU, and thus whether or not we will increasingly depart from the principle according to Article 10, that the Member States are responsible for implementation.

All in all, we can see that the role of a variety of actors is crucial when analyzing compliance with first pillar legislation in the EU. While the main activities take place at

²⁶ Commission of the European Communities (CEC), 25th Annual Report from the Commission on Monitoring the Application of Community Law, COM (2008) 777, Brussels, 2008.

²⁷ For the most recent statistics, please consult:

http://ec.europa.eu/community_law/directives/directives_communication_en.htm.

²⁸ Majone, G., The Credibility Crisis of Community Regulation, *Journal of Common Market Studies*, 2000. 38(2), p.273-302.

²⁹ See Groenleer, M., Kaeding, M. and Versluis E., Regulatory Governance through EU Agencies? The Role of the European Agencies for Maritime and Aviation Safety in the Implementation of European Transport Legislation, *Journal of European Public Policy*, (forthcoming).

the Member State level, we certainly must not lose sight of the Commission and agencies, and the strong role they play in the extent to which Community policies are actually complied with. Before analyzing what lessons we can learn to improve compliance in the third pillar from the role played by these types of actors in the first pillar, we must point to certain difficulties related to EU level oversight functions.

2.3 Difficulties related to EU level oversight functions

We are aware of certain obstacles and pitfalls in the functioning of the Commission as the ‘guardian of the Treaties’ that can potentially lead to valuable lessons for improving monitoring mechanisms.³⁰ The extensive monitoring and sanctioning powers accorded to the European Commission (in co-operation with the ECJ) in its guardian role requires careful evaluation and its impact should not be overestimated. The main lesson to be learned from the experience of the Commission in this role is that its impact on compliance is not straightforward and depends heavily on the availability of, and access to information by the Commission.

The main reason for questioning the Commission’s capabilities as guardian lies in the fact that, when analyzing in detail those infringement procedures that have been initiated, the Commission demonstrates a strong ‘bias’ towards only one specific policy instrument and only one part of the implementation process. Of the three binding European policy instruments used in the first pillar, around 85% of the annual 1,500 infringement procedures are focused on directives (while of the 9,000 legislative measures in force as at 2007, only about 2,000 are directives).

Furthermore, within this bias towards directives, a bias towards the transposition of these directives also appears. The Commission can initiate the procedure for four types of infringements – the absence of notification measures for directives,³¹ incomplete and incorrect transposition of directives, improper application of directives, and violations of Treaty provisions, regulations and decisions. However, in practice, about 70% of all procedures are initiated for transposition-related infringements (see Table 2).

³⁰ Versluis, E., *The Achilles’ heel of European regulation: national administrative styles and the Commission’s neglect of practical implementation*, ed Joachim, J., Reinalda, B. and Verbeek, B., *International Organizations and Implementation. Enforcers, managers, authorities?*, London, Routledge, 2008, p. 120 - 134.

³¹ Member States are obliged to notify the Commission of their transposition measures. An integrated system of electronic notification has been in existence since May 2004.

Table 2: Infringement procedures in 2006 per type of violation³²

PHASE IN INFRINGEMENT PROCEDURE	NUMBER	REASON FOR STARTING THE INFRINGEMENT PROCEDURE			
		Absence of Notification of Transposition of Directives	Incomplete / Incorrect Transposition of Directives	Improper Application of Directives	Violations Regarding Treaties, Regulations and Decisions
Letter of Formal Notice	1,536	905	205	201	225
Reasoned Opinion	680	371	85	112	112
Referral to Court	189	121	19	27	22

As stated, this ‘bias’ is heavily influenced by the availability of, and access to information by the Commission. Most information is available on transposition matters, while issues of practical implementation are less visible to the civil servants working in Brussels. While problems in the practical implementation stage seem to be serious, an ‘information deficit’ weakens the applicability of a coercive instrument in actually solving these problems. The Commission is therefore successful as a guardian in ensuring transposition of directives, but this tool is less applicable in ensuring practical implementation of all policy instruments. We can hardly blame the Commission for this: how can this relatively small organization have the necessary ‘eyes and ears’ in all 27 Member States, so as to be able to witness what happens in terms of practical implementation? The Commission does not have officials working in the Member States with the aim of locating infringements. It is therefore highly dependent on third parties to notify it about any shortcomings, but third parties are not equally active in all Member States or in all policy sectors.³³ Yet, it is crucial to the establishment of well-functioning oversight mechanisms at the EU level, is the need to ensure secure and reliable access to the necessary information.

Next to this more ‘traditional’ oversight function at the EU level, we observe a new trend in the first pillar: a more developed role for EU agencies as oversight mechanisms. While there is currently a period of reflection on the role of agencies in the EU regulatory landscape – and thus the future for agencies is still uncertain –there are already several

³² The year 2006 is no exception. Looking back to the numbers since 2000, we see similar percentages, based on Commission of the European Communities (CEC), 25th Annual Report from the Commission on Monitoring the Application of Community Law, COM(2008) 777, Brussels, 2008.

³³ To give one example: environmental NGOs are, in particular, very active in informing the Commission about situations of non-compliance. We thus see a relatively high proportion of infringement procedures in this sector. This is not to say, however, that compliance is a larger problem in this sector compared to other sectors: we can only conclude that environmental NGOs are more active compared to similar parties in other sectors. In other words, we have to be careful in drawing conclusions based solely on the infringement statistics of the Commission.

agencies with ‘compliance-related tasks’.³⁴ The most illustrative are the European Maritime Safety Agency (EMSA) and the European Aviation Safety Agency (EASA), which have their own inspectors who can visit inspection authorities in the Member States to determine the extent to which they sufficiently ensure compliance.³⁵

The EASA has, as its primary role, to ensure high and uniform levels of civil aviation safety in Europe. To this end, it aims to achieve standardization of national aviation safety legislation. The main instrument EASA has at its disposal to achieve standardization is via inspections of Member States. These inspections take the form of ‘inspections of the inspectorates’; in other words, EASA officials visit national aviation authorities to check how they are organizing their own inspections and thus how they ensure compliance.

Different Member States respond rather differently to this new trend. There has emerged a new European ‘interference’ in a domain traditionally organized at the domestic level, and national sensitivities and fears of losing power are visible. Member States with a long inspection tradition in the field of aviation safety (for example, Germany, France, and the Netherlands) in particular respond in a reserved manner.

Some of the new Member States, however, respond very positively. Polish aviation inspectors in particular indicate that they benefit from the recommendations of EASA officials and that they learn a lot from these visits; they also indicate that the level of compliance increased considerably due to EASA involvement.³⁶ It thus seems that the appreciation for, and reaction to, EU level involvement in enforcement varies as between different Member States. If one were to consider that agencies in the field of police and judicial cooperation play a stronger role in enhancing compliance –for example, the Fundamental Rights Agency³⁷ – one needs to take into consideration that domestic responses to such European involvement differ.

On a more positive note, agencies could potentially play a stronger role than the Commission. While the Commission’s activities as the guardian of the Treaties are only directed towards Member States, and thus only indirectly at the non-compliant behaviour of private actors, agencies could directly interact with the regulated.

2.4 Compliance with legislation with evaluation and monitoring mechanisms

As expressed in the research questions, the focus of this report is on the compliance with the outcomes of evaluation and monitoring mechanisms. Compliance with legislation, as opposed to compliance with evaluation and monitoring mechanisms, must be regarded as,

³⁴ Commission of the European Communities (CEC), 25th Annual Report from the Commission on Monitoring the Application of Community Law, COM (2008) 777, Brussels, 2008, p. 135.

³⁵ Groenleer, M., Kaeding, M. and Versluis E., Regulatory Governance through EU Agencies? The Role of the European Agencies for Maritime and Aviation Safety in the Implementation of European Transport Legislation, *Journal of European Public Policy* (forthcoming).

³⁶ Versluis, E., The role of EU agencies in securing compliance in central and eastern Europe, NIG Annual Work Conference, 20-21 November 2008.

³⁷ It is important to realize that the Fundamental Rights Agency is a Community agency.

in principle, two distinct issues. Whilst maintaining this focus, we identify a link between compliance with legislation and compliance with evaluation and monitoring. Many evaluations in the third pillar check compliance with legislation.³⁸ Compliance with legislation must be regarded as a precondition for compliance with the outcome of evaluations. In other words, without compliance with legislation, compliance with outcomes of evaluation may not be expected.

Additionally, the Commission regards evaluation as the first stage in the preparation of infringement procedures.³⁹ The evaluation criteria used by the Commission in the third pillar were made in the context of the first pillar. Although it recognises that it may not bring a legal action before the Court itself, it mentions that ‘there is a possibility for Member States to refer to the Court an allegedly incorrect interpretation or application (i.e. also transposition) of the framework decision by another Member State. The exercise of this legal possibility requires a solid basis on facts, to which this report is meant to contribute.’⁴⁰ This justifies and requires that attention is given to compliance with legislation.⁴¹ We will therefore do so, particularly in the following two paragraphs.

2.5 Analyzing compliance problems in the first pillar

All levels of government – more or less regularly – are confronted with implementation difficulties. International regimes are thought to be particularly vulnerable to compliance problems, due to the absence of a ‘legitimate monopoly of force to bring about compliance’.⁴² What are the specificities of compliance problems in the EU’s first pillar? Most of the scholarly literature on this topic is concentrated on the difficulties arising from the fact that the EU uses directives as a policy tool.

These instruments – which only dictate the result to be achieved, leaving the form and method to the Member States – require a transposition process; a process that can bring about specific difficulties. In particular, the proper functioning of the Internal Market led to the recognition that uneven and delayed transposition can harm the competitiveness of the market. Since the late 1990s, scholars have analyzed the timeliness and correctness (with a variation between incomplete and incorrect) of the transposition of directives in Member States.

³⁸ ‘This report should enable first of all the Council to assess the extent to which the Member States have taken the necessary measures to comply with the framework decisions’; COM (2001) 771, p. 7, and also COM (2004) 230, p. 7.

³⁹ COM (2004) 346, p. 4. See also Stine Andersen, Sovereignty and the emergence of non-binding peer review within the EU: The non-binding peer review mechanism facilitates a transition towards full enforcement in the TFEU fills gap article on file with the authors.

⁴⁰ COM (2001) 771, p. 7.

⁴¹ See for instance regarding the setting up of a methodology Report on the practical operation of the methodology for a systematic and rigorous monitoring of compliance with the charter of fundamental rights, Brussels, 29.4.2009, COM(2009) 205 final.

⁴² Börzel, T.A., Why do states not obey the law?, ARENA, University of Oslo, 6 June, 2002, p. 1.

Explanations for transposition problems vary, amongst others, from the quality of directives, to the goodness of fit, the attitude towards the EU, and any veto points, or domestic coordination problems.⁴³ Transposition problems are relatively well managed by the European Commission (see above). In comparison to other international regimes, the EU has the most institutionalized and legalized enforcement system.⁴⁴ The European Commission as the ‘guardian of the Treaties’ seems to have a rather successful impact upon transposition results and the EU thus seems to be ‘exceedingly effective in combating detected violations’.⁴⁵

Even if transposition problems are relatively well managed, the application of the rules can still cause problems. First of all, there are many more policy instruments than directives alone.⁴⁶ Regulations and decisions also need to be put into practice. These instruments, as well as directives once properly transposed, need to be enforced by the regulators and – more importantly – they need to be complied with by the regulated. Empirical insight into the level of enforcement of, and compliance with, EU legislation is still relatively low. This is mainly due to the scarcity of data and the difficult and time-consuming nature of collecting such data.⁴⁷ The few available case studies show that even rules do not automatically lead to even practices. The Safety Data Sheets Directive, for example, is perfectly transposed by Member States, but the directive is hardly enforced by national inspectors and in many chemical companies – particularly the smaller and medium-sized enterprises – the safety data sheets required for dangerous chemical products are either missing, or of a low quality.⁴⁸ In other words, the law in the books does not automatically lead to law in action.

It is argued that, since the Central and Eastern European countries (CEEC) joined the EU in 2004 and 2007, compliance problems have become more relevant.⁴⁹ Even during the pre-enlargement period, the Commission decided to pay specific attention – in addition to the Copenhagen criteria – to the administrative and judicial structures needed for effective implementation (Madrid criteria, 1995). While compliance problems are certainly not limited to the new Central and Eastern European Member States, it is expected that such problems are more likely to occur there. Schimmelfennig and Sedelmeier state that the CEEC have a reputation of not ‘fully and reliably’ implementing

⁴³ For an overview of the many variables provided see Mastenbroek, E., *EU Compliance: Still a ‘Black Hole’?*, *Journal of European Public Policy*, 2005.

⁴⁴ Zürn, M. and Joerges, C., *Law and Governance in Postnational Europe, Compliance beyond the Nation-State*, Cambridge: Cambridge University Press, 2005.

⁴⁵ Tallberg, J., *Paths to Compliance: Enforcement, Management, and the European Union*, *International Organization*, 2002, p. 610.

⁴⁶ As counted by the European Commission in 2007, of the more than 9,000 legal measures in place, less than 2,000 are directives. See Commission of the European Communities, *A Europe of Results – Applying Community Law*, COM (2007) 502, 5 September, 2007, Brussels, p. 1.

⁴⁷ Treib, O., *Implementing and complying with EU governance outputs*, *Living Review European Governance*, No. 1, 2006. Available at <http://www.livingreviews.org/lreg-2006-1>.

⁴⁸ Versluis, E., *Even rules, uneven practices: Opening the ‘Black Box’ of EU law in action*, *West European Politics*, 2007.

⁴⁹ Schimmelfennig, F. and Sedelmeier, U., *Conclusions: The impact of the EU on the accession countries*, ed. Schimmelfennig, F. and Sedelmeier, U., *The Europeanization of Central and Eastern Europe*, Ithaca, NY: Cornell University Press, 2005.

the transposed rules, and these scholars are rather pessimistic about the future. After their accession, the external incentives to comply disappeared, and ‘the absence of these incentives should significantly slow down or even halt the implementation process’.⁵⁰

The main reasons provided for the lack of practical implementation are administrative capacity limitations and weak post-communist societal mobilization.⁵¹ This particularly relates to the weak institutional structures and scarcity of technical expertise in both enforcement organizations (including an absence of skilled inspectors) and in the judiciary, which obstruct successful compliance. Since the Commission as the guardian of the Treaties to a large extent depends on complaints and notifications in order to detect infringement cases,⁵² the weak civil society in most CEEC negatively influences the possibility for detecting non-compliance.

In addition, Falkner and Treib⁵³ identify the negative influence of the common practice of literally transposing EU directives in most CEEC, as this implies a lack of adaptation to local situations. Hille and Knill⁵⁴ demonstrate that, in particular, a country’s administrative strength – the quality and effectiveness of its bureaucracy – influences the implementation performance. In addition to these more ‘managerial’ reasons, Goetz also distinguishes a more ‘rational’ explanation. Besides lacking capacity, we can also question the willingness to comply.⁵⁵ All in all, the CEEC seem to show good transposition records, but experience difficulties with the practical application of Community law.

Explaining compliance problems

When analyzing the academic literature on compliance difficulties, we observe that there are a variety of factors that seem to be relevant in explaining why countries do or do not comply with EU legislation. The Annex to this report provides an overview of the main factors identified in the relevant literature.

Compliance problems result from two sets of factors: those at the EU level and those at the Member State level. From the ‘compliance theories’ in the ‘International Relations’ (IR) literature that analyze why nations comply with international legislation, we know

⁵⁰ Schimmelfennig, F. and Sedelmeier, U., Conclusions: The impact of the EU on the accession countries, ed. Schimmelfennig, F. and Sedelmeier, U., The Europeanization of Central and Eastern Europe, Ithaca, NY: Cornell University Press, 2005, p. 226.

⁵¹ Falkner, G. and Treib, O., Three worlds of compliance or four? The EU-15 compared to new Member States, *Journal of Common Market Studies*, 2008.; Schimmelfennig, F. and Sedelmeier, U., Conclusions: The impact of the EU on the accession countries, ed. Schimmelfennig, F. and Sedelmeier, U., The Europeanization of Central and Eastern Europe, Ithaca, NY: Cornell University Press, 2005, p. 226.

⁵² This relates to about 40% of the cases; Commission of the European Communities (CEC), A Europe of Results – Applying Community Law, COM (2007) 502, 5 September 2007, Brussels.

⁵³ Falkner, G. and Treib, O., Three worlds of compliance or four? The EU-15 compared to new Member States, *Journal of Common Market Studies*, 2008.

⁵⁴ Hille, P. and Knill, C., ‘It’s the bureaucracy, stupid’ The implementation of the *Acquis Communautaire* in EU candidate countries, 1999-2003, *European Union Politics*, 2006.

⁵⁵ Goetz, K., *The new Member States and the EU: Responding to Europe*, Ed Bulmer, S. and Lequesne, C., *The Member States of the European Union*, Oxford: Oxford University Press, 2005, p. 275 - 276.

that there are three main schools of thinking about compliance – rationalism, management and constructivism – that each operate from different assumptions. Each of the three approaches stresses different reasons for the existence of compliance problems and thus also assumes different types of solutions.

Rationalism dominated IR thinking regarding compliance in the 1980s and is anchored in the political economy tradition of game theory and collective action theory.⁵⁶ States are perceived as making rational choices – to decide whether or not to comply – on the basis of cost/benefit calculations. This perspective thus stresses intentional mechanisms and labels non-compliance as either a preference or as opposition.

What options does the EU have to address compliance problems according to a rationalist account? The answer is straightforward: compliance requires enforcement. When there is no effective system to detect and respond to violations or infringements, actors will not comply; thus only a coercive strategy of monitoring and sanctioning will induce compliance.

AS opposed to rationalism, the managerial perspective departs from the idea that States are generally willing to comply with international rules and that overall compliance rates are relatively good.⁵⁷ When States do not comply with international legislation, this is rather the result of capacity limitations or rule ambiguity. In order to reduce problems, non-compliance should be ‘managed’ instead of sanctioned. Managerialism is rather explicit in providing solutions to the compliance puzzle; solutions are to have a cooperative and problem-solving approach based on capacity building (for example, funding or sharing of best practices) and rule interpretation (for example, guidelines, EU wide inspection criteria).

While the rationalist and management perspectives are often presented as the two main ends of the spectrum when theorizing compliance, since the late 1990s we see the rise of an alternative approach that is less concerned with thinking in terms of intentional or unintentional mechanisms, but instead is more concentrated on a normative analysis of compliance.⁵⁸ According to constructivists, States are persuaded to comply with international law as their preferences change due to the socialization or internalization of shared norms.⁵⁹ Non-State actors – for example, NGOs, relevant communities, churches, universities or research centres, media, trade unions – are particularly considered to play an important, catalytic role in generating pressure and thus stimulating social learning.

The most important lesson we can draw from the list of factors that ‘compliance literature’ generates, is that non-compliance can hardly ever be attributed to a single

⁵⁶ Koh, H.H., *Why Do Nations Obey International Law*, The Yale Law Journal, 1997.

⁵⁷ Chayes, A. and Chayes, A.H., *On Compliance*, International Organization, 1993.

⁵⁸ Raustiala, K. and Slaughter, A.M., *International Law, International Relations and Compliance*, Ed. Carlsnaes, W., Risse, T. and Simmons, B.A., *Handbook of International Relations*, London: Sage Publications, 2002.

⁵⁹ Checkel, J.T., *Why Comply? Social Learning and European Identity Change*, International Organization, 2001.

factor or explained by a single perspective.⁶⁰ Falkner et al,⁶¹ in a study of 29 infringement cases, showed that many different explanations –for example, deliberate opposition, administrative shortcomings, interpretation problems – are in place at the same time. The different compliance theories are not competing; the mechanisms are most effective when combined.⁶² As Underdal argues, the question is not which model is ‘true’, but more how much of the variance in compliance each model can explain.⁶³ Or, as Coleman and Doyle state, the question is not whether the various models on compliance are true, ‘but *when* they are true’.⁶⁴

2.6 Analyzing compliance problems in the third pillar

The three general research questions seek to find the causes of non-compliance with evaluation and monitoring mechanisms in the third pillar. As expressed in 2.4, the first level is the one dealing with compliance factors with the legal instruments themselves. After all, compliance with evaluation and monitoring mechanisms is tested against the compliance with the legal instruments. It is therefore necessary to look at factors that influence compliance with legislation.

Compliance factors regarding legislation

This section on compliance factors can be divided, partly chronologically, into various categories. First, attention is paid to the drafting and negotiating process. Then the result of that process is looked at: the adopted legal instrument. We will deal with the vague or precise norm setting and the clarity of the obligations deriving from the instrument. In addition, other factors that complicate compliance will be identified, as well as the use of non-defined principles.

1. Drafting and negotiating process

Regarding the drafting and negotiating process, many existential questions arise. Was the legal instrument preceded by an impact assessment? Why was the legal instrument necessary? What are the precise problems for which this is a solution? What we can see from these questions is that problem analyses and impact assessment reports are

⁶⁰ Neyer, J. and Zürn, M., Compliance in Comparative Perspective. The EU and Other International Institutions, InIIS-Arbeitspapier, nr. 23/01, Universität Bremen, 2001.

⁶¹ Falkner, G., Hartlapp, M., Leiber, S. and Treib, O., Non-Compliance with EU Directives in the Member States: Opposition through the Backdoor?, West European Politics, 2004.

⁶² Raustiala, K. and Slaughter, A.M., International Law, International Relations and Compliance, Ed. Carlsnaes, W., Risse, T. and Simmons, B.A., Handbook of International Relations, London: Sage Publications, 2002; Checkel, J.T., Why Comply? Constructivism, Social Norms and the Study of International Institutions, ARENA Working Papers, 99/24, Oslo: Advanced Research on the Europeanisation of the Nation-State, 1999; Börzel, T.A., ‘Why do states not obey the law?’, ARENA, University of Oslo. 6 June, 2002.

⁶³ Underdal, A., Explaining Compliance and Defection: Three Models, European Journal of International Relations, 1998.

⁶⁴ Coleman, K.P. and Doyle, M.W., Introduction: Expanding Norms, Lagging Compliance, Ed. Luck, E.C. and Doyle, M.W., International Law and Organization, New York, Rowman & Littlefield, 2004, p. 10 (emphasis in original).

generally, albeit not completely, absent.⁶⁵ It seems that they exist for Commission proposals, but not for Member State initiatives. Impact assessment reports may contain assessments of the (practical, not legal) advantages and disadvantages of certain options. The stated explanations or rationales for the draft legislation are generally poor.⁶⁶ Particularly with regard to the criminalization of conduct and the method of combating crime, the absence of any criminological research is striking. One would have expected some insight in the forms of undesired conduct, to which the drafting of legal instruments would correspond.

In the context of the drafting process, it is also relevant to look at the stakeholders. What was the goal of the negotiating parties? Did they want to achieve something or were they more concerned about damage control? Were Framework Decisions adopted to give the Presidency its own legal instrument? These are questions that determine the ownership of the legislation and will later influence findings about compliance with evaluation.

The decision making process is directed more towards making use of the political momentum than towards the best balanced legislation. Delays in decision making are often regarded as failures. See, for example, the Commission monitoring of the adoption of measures scheduled for 2007 under The Hague Programme: 'The general overall assessment is rather unsatisfactory. A significant number of actions envisaged (...) had to be abandoned or delayed (...) The 2007 report reveals a lower rate of achievement (38% of measures achieved) compared to 2006 (53% achieved), with a substantial increase in actions that had to be delayed: 41 % compared to 27% in 2006.'⁶⁷

It is obvious that the Member States, the Commission and the Council very much participate in the drafting process of legislation, but that there is a general absence of practitioners and civil society involved in the process of drafting new legislation.⁶⁸ There is no direct evidence that the judiciary, the prosecution, the bar or the police are consulted in a structural manner. This is to be regretted, since it would broaden the perspective and the expertise of these fora. Additionally, it would certainly improve the quality of legislation. There is, of course, a dark number in the sense that it may be that one of the Member States or institutions participating in the drafting did consult practitioners.

If stakeholders are consulted, they do not include the criminal bar. Criminological research is not conducted; the impact assessment seems to be made by lawyers only and

⁶⁵ See, for example, Report by the Friends of the Presidency on the technical modalities to implement the principle of availability, Brussels, 10 November 2005, 13558/1/05 REV 1.

⁶⁶ Voorstel voor Kaderbesluit van de Raad betreffende de uitwisseling van informatie volgens het beschikbaarheidsbeginsel, Brussel, 12.10.2005 COM(2005) 490 definitief; Proposal for a Directive of the European Parliament and of the Council providing for sanctions against employers of illegally staying third-country nationals, Impact assessment, Brussels, 16.5.2007 SEC (2007) 603. Proposal for a Council Framework Decision amending Framework Decision 2002/475/JHA on combating terrorism, Impact assessment, Brussels, 13 November 2007 14960/07 ADD 1.

⁶⁷ Communication from the Commission to the Council and the European Parliament, Report on Implementation of The Hague Programme for 2007, Brussels, 2.7.2008, COM (2008) 373 final, p. 2.

⁶⁸ This is so, even though it has been proposed by the Commission. See Communication from the Commission to the Council and the European Parliament, Evaluation of EU Policies on Freedom, Security and Justice, COM (2006) 332 final, Brussels, 28.6.2006, p. 8.

generally from their desks in Brussels, more on the basis of strategic reasoning than on the basis of evidence-based research.⁶⁹ They emphasize considerations of a policy and speculative strategic nature.⁷⁰ There is a general tendency to look at crime with normative concepts. Empirical evidence as to why existing legislation is unsatisfactory is not collected or not produced.

2. The adopted legal instrument: vague or precise norm setting and the clarity of the obligations deriving from the instrument

Whilst many provisions of third pillar legislation are very clear, others suffer from vagueness of the normative standard in the legal instrument. An example of a relatively non-binding norm is taken from the definitions of offences - Article 2, paragraph 1, Framework Decision 2005/222 on Attacks against Information Systems, reads:

‘Each Member State shall take the necessary measures to ensure that the intentional access without right to the whole or any part of an information system is punishable as a criminal offence, at least for cases which are not minor.’

The probability that Member States will significantly differ is extremely high: what are ‘the necessary measures’, what is ‘intentional’, what is ‘without right’, and which cases are ‘not minor’? Here, the approximating effect is practically non-existent. This ambiguity has the potential to give rise to imprecision about what implementation in this case really means, and this will come back at the level of evaluation and monitoring and may lead to disputes between Member States and an evaluation team.

Another example is Framework Decision 2004/757 on Illicit Drug Trafficking. This Framework Decision imposes the obligation to set at least a penalty of a given length as the maximum penalty for the offence.⁷¹ The degree of freedom that a Member State has in approaching this is significant. It can opt to establish a longer penalty. If the Member State already has a longer penalty, it need do nothing. In sum, the quality of legislation must be improved.⁷²

Is the new set of rules exclusive? Does it replace the previously existing system or does the new legal instrument create an additional layer of legislation? Do alternatives exist? In this respect, one may refer to Framework Decision 2002/584 on the European Arrest Warrant. This is still the only Framework Decision implemented in all Member States.

⁶⁹ It is questionable whether this will be changed by the system announced by the Commission of Extended Impact Assessment. See Justice Forum Communication, par. 12. More promising is the Commission Decision 2006/581/EC of 7 August 2006, setting up a group of experts on the policy needs for data on crime and criminal justice, OJ 2006, L 234/29.

⁷⁰ Towards a general policy on the fight against cyber crime; Summary of the Impact assessment, Brussels, 30 May 2007, 10089/07 ADD 1 and 2.

⁷¹ Article 4, paragraph 1, Framework Decision 2004/757 on Illicit Drug Trafficking: ‘Each Member State shall take the necessary measures to ensure that the offences referred to in Article 2 are punishable by criminal penalties of a maximum of at least between one and three years imprisonment’.

⁷² Report by the Future Group (Justice), Brussels, 22 July 2008 (23.07), 11962/08.

One of the factors contributing to full implementation is that the European Arrest Warrant replaces the existing system of extradition relationships.

In other words, a Member State not complying with the obligations to implement could no longer cooperate with other Member States. This may be regarded as a tremendous incentive to comply. Most other new legal instruments are additional to existing legal instruments and do not replace them. It is submitted that instruments that replace the existing mechanisms in international cooperation are implemented and complied with to a much higher degree than other instruments of a less exclusive character.

What are the legal consequences of non-compliance or non-implementation? At least in the third pillar, there are essentially none. This is due both to the absence of the enforcement obligation in the third pillar (although the Pupino case has raised some of the consequences) and the fact that the Commission may not bring a case against a Member State. The possible results of such a situation in which compliance cannot be enforced are self-evident; even in the situation of the exclusive mechanism of the European Arrest Warrant, it took the Member States more than a year after the stipulated period until all had implemented it.⁷³

The current regime also does not provide for any sanctions for Member States who do not wish to cooperate with another Member State, due to a lack of trust in that Member State's respect for the rule of law. Given that fact, it is interesting to see that Member States have refrained from directly using deficiencies in the respect for the rule of law in a Member State as a reason to refuse cooperation. One can only speculate as to the reasons for this. No Member State has ever brought a complaint against another Member State for not respecting third pillar legislation or, more generally, for not respecting the rule of law. Similarly, on the level of compliance with the findings of peer evaluation, there is no sanction when a Member State does not comply with its findings.⁷⁴

The second reason, apart from vague norm setting, for the meagre influence of integration and harmonization, is the use of 23 legal languages. By definition, the use of various language versions leads to differences in some legal instruments. In the European 'Tower of Babel', every Member State interprets European law in its own (legal) language, and talks at cross purposes with other States. The language factor reappears at the level of evaluation and monitoring.

⁷³ Member States should have implemented it by 31 December 2003. Twenty-four Member States implemented it with up to 8 months delay and then had to wait for Italy as the last State. Finally, on 12 May 2005, the Framework Decision took effect. See Report from the Commission based on Article 34 of the Council Framework Decision of 13 June 2002 on the European Arrest Warrant and the surrender procedures between Member States, Brussels, 23.02.2005, COM (2005) 63 final.

⁷⁴ See, for example, Letter of the Minister of Justice of the Netherlands to Parliament on the follow up to the findings of the team evaluating the implementation of the European Arrest Warrant in the Netherlands. The Minister announced that most of the recommendations will be complied with, but also mentions that there are some that will not be complied with, the most prominent example being the deletion of Article 11 of the Surrender Act, stipulating that surrender will not be granted if there is fear that it will lead to a flagrant denial of justice. See Parliamentary Documents Second Chamber, 2008/2009, 23490, nr. 545, p. 4-5.

The fact that implementation is inevitable when it comes to introducing criminal offences is the third reason for the divergence of implementation and enforcement in Member States. Common standards will be translated and incorporated into a national context. This is the cumulative effect of vague standard setting in various languages and is conducive to an interpretation that will change as little as possible in national law. National law enforcement officials will interpret the (new) rules with national eyes and as part of an entire national system. This explains why some rules that - on paper - ought to change standards in a specific country, may in practice not result in any noticeable change. Harmonization that is limited to common European rules may, therefore, merely lead to formal harmonization.⁷⁵ Additionally, a Member State that comes off the worst in the negotiation process can still retaliate afterwards by interpreting the implementation obligations in an extremely nationalistic manner.

The fourth reason is found in the incidental guidance provided by the Court. The actual role of the Court in the interpretation of Framework Decisions in preliminary questions is gradually taking on more shape, but still remains fragmented, unsystematic and incidental. The fact remains that not all Member States give the Court jurisdiction in the same way. There are countries that give all national courts the competence for preliminary references, but there are others that restrict this to their highest court. Additionally, national courts are rather reluctant to refer questions to the Court.⁷⁶ This may have to do with the autonomous interpretation the Court will give to legislation. The vaguer the norms are the more creative the Court must be and the more discretion it has.

The position of the ECJ seems to go completely unnoticed and unaffected in all studies on compliance. However, although the Court is the highest authority on the interpretation of the Treaties and the secondary legislation, it also uses its position to bring coherence to fragmentary legislation, thus contributing to further compliance with the long term ideal of European integration. In the absence of infringement proceedings, compliance with legislation could be stimulated by more preliminary references.

The fifth reason lies in the absence of a uniform view of the role of criminal law in the Union. There is no grand design on the relationship between Union law and criminal law. This might even be the most important factor creating the gap between the common instrument and the law in practice. This strengthens the idea that compliance is only about compliance with a specific legal instrument, and not with a larger whole, such as the rule of law.⁷⁷ Legislative initiatives are taken on single issues, without analysing the consequences in a broader context. In that sense, it is relevant that, in the choices of the topics to legislate, the approach is fragmentary, incoherent and inconsistent. This characteristic reappears in relation to compliance with evaluation and monitoring mechanisms. Because there is no coherent framework to consider, the scope of evaluation

⁷⁵ A more cynical view would suggest that ‘window-dressing’ takes place.

⁷⁶ We have not been able to conclude that vague norm setting in the third pillar did lead to an increase of the number of preliminary references.

⁷⁷ At first sight, it might be easy to regard respect for the rule of law as the key guarantee in the area of freedom, security and justice. However, the EU has not formulated any specific policy to that end.

is limited to looking at compliance with, and implementation of a specific legal instrument.

3. The use of non-defined, although leading, principles

Vague standard setting is general: both regarding the role of criminal law in the EU, as well as the applicable principles. Mutual recognition has not been defined, despite its 'corner stone' status.⁷⁸ What mutual recognition means exactly leads to discussions and disputes, both at the level of complying with legislation as well as with the findings of evaluation mechanisms. Likewise, the area of freedom, security and justice, the principles of subsidiarity and proportionality, and other fundamental aspects of the national criminal justice system, are not well defined. The absence of definitions creates more freedom for Member States to go their own way and to give things their own interpretation. This is not limited to compliance with the binding norm, but has similar consequences at the stage of compliance with evaluation findings.

Although the internal market and the area of freedom, security and justice have been merged in the Court's case law, its consequences for the respective roles of Member States and individuals are still a work in (slow) progress. The inclusion of non-economic fields into the Union legal order requires certain adjustments of long-standing principles, due to a changed perspective in which the interests of EU citizens are central. EU citizens have their own rights, which may come into conflict with the rights of the Member States. The opposite perspectives of those who enjoy rights in this area require that a balance be found. While the internal market grants rights to citizens, the area of freedom, security and justice predominantly grants rights to the Member States.

To apply the principles developed in the internal market in exactly the same manner in the latter context will lead to friction. The recognition of rights to EU nationals outside of the economic sphere will require the formulation of additional, or new general principles of Union law. This may lead to principles that are much more fundamentally rights-oriented, rather than oriented towards the economic well being of the Union. Such principles may relate, for example, to the relationship between the principle of proportionality and the principle of mutual recognition in surrender proceedings, or to the meaning of the rights of the defence in mutual recognition. It is useful to recall the words of the Court in *van Gend en Loos*, when it dealt with the new legal order that had come into being: 'The new legal order is comprised not only of the Member States; their nationals constitute subjects of this new order too'.⁷⁹

The principle of availability serves the exclusive interests of law enforcement authorities. The exchange of privacy-sensitive data upon the basis of availability requires surveillance of data protection rules. The fact that an increasing number of mechanisms are constructed in which respect for data protection rules is tested by those authorities that have an interest in gaining access to the information, is not conducive to an objective

⁷⁸ See further European Criminal Law 2009, p. 330 - 354.

⁷⁹ 5 February 1963, Case 26-62, *Van Gend en Loos* [1963] ECR I.

approach. In a Union legal order that is striving for the free availability of all the information in the possession of Member State authorities, data protection is essential.

Some authors argue that a real area of freedom, security and justice can only be achieved by more harmonization in the field of the rights of citizens.⁸⁰ It is certainly true that mutual recognition is easier when the legal systems are closer to each other. Is it the case that, after a decade of primary attention on the harmonization of criminal law (1993-2002), there has been a decade (since Tampere 1999) where the main focus of attention is on mutual recognition? It may actually be the right time for harmonization again. Harmonization and mutual recognition should be regarded as complementary, not as mutually excluding alternatives.

Mutual trust has been declared as a dogma. It lies at the heart of the principle of mutual recognition. It presupposes common values,⁸¹ the existence of which, however, has never been tested and evaluated. The existence of grounds of refusal in the legal instruments gives Member States the possibility to protect the interests of the individual and the Member State itself in their own way. Whilst first pillar legal instruments on mutual recognition contain rules on the allocation of jurisdiction, such jurisdictional rules are absent in the third pillar. Third pillar instruments in criminal law combine grounds for refusal with the absence of an allocation of the jurisdiction of the Member States.⁸²

Most grounds for refusal, however, relate to cases over which there is multiple jurisdiction - see, for example, Articles 3 and 4 of the Framework Decision 2002/584 European Arrest Warrant. These grounds relate either to the executing Member State having jurisdiction over the offence itself, or that the request relates to a national of the Member State.⁸³ Both grounds for refusal and jurisdiction require Member States to formulate their own opinion on the prosecution initiated by another Member State. This is more conducive to undermining mutual trust than to creating it. It is recommended that grounds for refusal will be limited and that extraterritorial jurisdiction in other Member States will either be limited or allocated.

Compliance factors in the Member States

The starting point is that the primary responsibility to comply with EU law lies with the Member States. In the field of substantive criminal law, direct applicability of Union law, without national transposition, is prohibited.⁸⁴ Member States must therefore implement

⁸⁰ Manuel Monteiro Guedes Valente, A construção de um espaço penal europeu sob o primado dos direitos fundamentais: sonho ou realidade?, in: Lorenzo M. Bujosa Vadell, Hacia un verdadero Espacio Judicial Europeo, Granada 2008, p. 3-20. This also seems the position taken by the De Schutter.

⁸¹ Olivier De Schutter, The Role of Evaluation in Experimentalist Governance: Learning by Monitoring in the Establishment of the Area of Freedom, Security and Justice, Faculté de Droit de L'Université de Louvain, November 2008, p. 2.

⁸² This is different in civil law. Article 34 Regulation 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters, contains a much more limited and clearer catalogue of grounds for refusal. See further Klip, European Criminal Law, p. 330-354.

⁸³ With the exception of Article 3 sub 3 (minor) and Article 4 sub 1.

⁸⁴ See Klip, European Criminal Law,, p. 204-207.

criminal legislation, in order to meet their obligations under the Treaties. This leaves us with the questions of what techniques the Member States use in practice when implementing the obligations to criminalize, and whether some conclusions can be drawn regarding implementation and compliance.

Additionally, the question also arises as to whether the implementation techniques predict something about compliance with evaluation outcomes. The legislative practice of Member States in implementation shows a wide variety of techniques, even within the one Member State. Union law leaves Member States free in the use of these techniques. It is the result that counts, the standard phrase in Framework Decisions being: ‘Each Member State shall take the necessary measures to ensure that the following intentional conduct is punishable.’

Using a ‘copy and paste’ method, the Member State copies the wording of the prohibited conduct into a provision of national criminal law. This seems to be a very safe method when a Member State wishes to live up to its obligations. It makes it quite easy to apply national legislation in conformity with the Union act. However, the dangers of this method lie in the use of terminology from the Union act that may not be fully compatible with the national criminal justice system and could thus lead to distortion.

Another (at first glance) ‘safe’ method is the reference technique.⁸⁵ Here, a Member State will adopt national criminal legislation which states: ‘It is shall be a criminal offence to commit the conduct as described in Article xy of Framework Decision YZ.’ This method may be problematical in some national systems, because the Union instrument may use terminology other than used to date in the criminal justice system, thus perhaps leading to confusion. However, unlike the copy and paste method, the reference method may not oblige Member States to amend its legislation if there is an amendment of the European instrument. Naturally, this applies only if the amendment relates to the provisions referred to in the national criminal law alone. This method of implementation incorporates any changes to the Union legal instruments automatically. From a legislative perspective, this is mainly apparent with technical regulations, to which frequent changes can be expected.

The reference technique may come into conflict with the *lex certa* principle, if it must be understood to require that a national written norm fully describes the prohibited conduct. It may be expected that Member States using the copy and paste method or the reference method will be inclined to consider their implementation as compliant by definition, which may influence their position in the evaluation process.

A more common technique is that of translation. By this method, the Member State translates the obligations deriving from the European instrument into its own words. This allows the Member State to maintain the coherence of its national criminal justice system. However, even when the Union legal instrument is adopted in every legal language, the

⁸⁵ The obligation to refer to the Union legal instrument in the implementing legislation should be regarded as a way of recognizing its European roots and corresponding connotations, not as an obligation to use direct reference in the criminalization process. See Article 291, paragraph 4 TFEU.

terms under Union law may not fully correspond to the terminology used in national law. This may be explained by the fact that a number of language versions cover more than one criminal justice system (Dutch, English, French, German, Greek and Swedish). The translation technique is generally used in the area of judicial and police cooperation. Particularly since the introduction of mutual recognition of cooperation in criminal matters, Member States have followed the road of legislating in specific acts the obligations from the EU instrument.

The last method found is not necessarily one of legislation, but one of application: the judicial European interpretation of the elements of the national definition of a crime. An example of the latter is where a national criminal court interprets an existing provision of the national criminal code in the light of the current Union legislation. In a situation where the national provision is the product of the implementation of European obligations, this must already be done in the form of an interpretation that is in conformity with the Directive and Framework Decision.

Additionally, there are also situations in which the significance of a term from the description of the offence is further defined by the underlying Union law. These can either be very broad terms, such as ‘unauthorized entry or residence’, ‘unlawful’, or very specific terms, such as ‘waste products’. National criminal courts will be legitimized to interpret the national terminology in a Union-minded manner, without further implementation. The limits of interpretation in conformity with Directives and Framework Decisions lie in the principle of legality.⁸⁶

Another issue is where, in the national criminal legislation, the implementing legislation should find its place. It is apparent that Union law does not point Member States in any specific direction. It is a matter left entirely to the Member States. In the process of implementing legislation, the national court must jump through all manner of hoops to maintain the existing system in place. Choices must be made. Will transposition be in the general Penal Code, or in a separate act? Member States may have all kinds of motives to incorporate as much as possible either into the existing system involving a general code, or into special criminal acts. The incorporation into the Penal Code may be more conducive to maintain the coherency of the national criminal justice system. However, the maintenance of this coherency could also provide a good reason to do the opposite. Implementation of a Union norm can have a disruptive effect at the national level.

On the other hand, adoption of a separate act may provide more recognition to the Union origins of this piece of legislation. Occasionally, one may have doubts as to whether the adoption of a separate act is not influenced by the desire to give its existence as little attention as possible. The lack of any case law on the compliance of the Member States with the obligation to criminalize behaviour originates from the fact that the Court will not have jurisdiction over such cases until the entry into force of the Treaty of Lisbon.

⁸⁶ ECJ, 3 May 2005, *criminal proceedings against Silvio Berlusconi and others*, C-387/02, C-391/02 and C-403/02 [2005] ECR I-3565.

One of the consequences of the fact that mutual trust is supposed to already exist and may not be put in doubt, is that an individual must seek his/her legal remedies in the Member State that issues the instrument of mutual recognition. An individual, not convinced of the fact that the issuing Member State deserves trust, will, in principle, not be able to complain about that in the executing Member State. However, there are indirect and alternative ways to do so. In the field of cooperation, Member States interpret grounds for refusal extensively, not restrictively, although the ECJ states that they should be interpreted restrictively.

Member States seem to join in the rhetoric in Brussels when adopting legal instruments and return to their own interests back home. They seem to neglect the fact that the stage of European integration has seriously limited the possibilities of a specific criminal policy. The question must be asked as to whether the absence of a common European criminal policy, given the level of integration, is not one of the factors contributing to a lack of mutual trust. Criminal policy is still formulated on a national level, not on a European level. This leads to the suggestion that issues relating to criminal law belong to the competence of the Member State.

2.7 Lessons learned

This second chapter has set the scene by providing an insight into compliance in the first and third pillars and the main problems associated with this. Furthermore, it provided an insight into the main oversight functions at the EU level, and the categories of factors that can be identified when analyzing compliance problems (see also the Annex to this report).

This overview shows us that compliance problems come in various forms and degrees. In the first pillar, where the EU works to a large extent with binding legislation, we can identify both transposition and practical implementation difficulties. While the Commission as ‘guardian of the Treaties’ has a relatively strong tool at hand to combat transposition problems in the first pillar, the analysis above indicates that this tool is less effective when it comes to combating practical implementation problems.

Four main lessons can be learned from this are as follows:

1. In order to ensure better compliance with first pillar legislation, *we need of more detailed information about the specific compliance problems at hand*. Currently, it is very difficult for the Commission as guardian to initiate the infringement procedure for problems related to practical implementation, due to an information deficit. This deficit figures prominently not only in the first pillar, however; limited access to information is also problematical in the third pillar. **Crucial to the establishment of well-functioning evaluation and monitoring mechanisms at the EU level, is the need to ensure secure access to the necessary information about what specific compliance problems exist in the Member**

States, and by what factors they are influenced, in order to adjust the evaluation and monitoring mechanisms accordingly.

2. We need to *improve our level of awareness of the importance of domestic, street-level, factors*. The overview of factors that explain compliance problems (see also the Annex to this report) demonstrates how crucial it is to realize that there is not one single factor that explains the lack of compliance. In different situations, and for different Member States, different factors are at play. **Crucial to the creation of well-functioning evaluation and monitoring mechanisms at the EU level, is the realization that it is impossible to find ‘one perfect tool’ that will equally trigger all Member States to comply, and that we need to become aware of the fact that domestic circumstances matter.** The multitude of identified factors also demonstrates that we can categorise different types of factors (management, rationalist and constructivist) that all need a different focus in terms of solutions to overcome compliance problems.
3. Vague norm setting, as a result of compromises, in combination with the absence of sanctions for non-implementation in the third pillar, undermine ownership of legislation and may explain shortcomings in compliance. The intergovernmental structure of the third pillar supports the view that the formulation of criminal law is still part of the full sovereignty of the Member States. **The broader context in which specific legal instruments in the field of criminal law are adopted should be made more explicit.** It should be clear what the thoughts behind a single instrument are, and how it fits into the body of other EU and national legal instruments. Looking at a legal instrument without taking its context into consideration may be counter-productive.
4. Whether a Member State has complied with third pillar legislation or not can, thus far, not be tested by the highest authority on these issues, the ECJ. Differences of opinion as to whether there is compliance typically do not end and may continue at the stage of evaluations. **The absence of infringement procedures in the third pillar undermines both compliance with implementation, as well as compliance with the findings of evaluation mechanisms.**

3. Setting the scene: monitoring and evaluation mechanisms

Having provided the necessary context regarding factors that either hinder or stimulate compliance with first and third pillar legislation, we now turn to a more specific question: how to improve compliance with monitoring and evaluation mechanisms? This chapter approaches this question by first of all studying the functioning and effectiveness of the peer review mechanism at work in the Organization for Economic Cooperation and Development (OECD), and then examining the use of such mechanisms within the framework of EU first-pillar cooperation.

Both discussions will provide the basis for some general suggestions, provided in chapter 4, as to how we could enhance the effectiveness of evaluation mechanisms. Also in that chapter, the existing mechanisms applicable in the third pillar are identified and described. In chapter 5, they will be analysed further. Additionally, we will examine what lessons emerge from the analysis and what recommendations can be made.

3.1 Learning from the peer review mechanism used in the OECD

We are in the fortunate situation that an analysis of the monitoring and evaluation mechanisms in the EU does not have to start from scratch. There are a number of international organizations that have a long tradition of using the type of evaluation mechanisms that have increasingly been employed in the EU, and to which the Union has in fact resorted for inspiration. This section will provide a short analysis of both the functioning as well as effectiveness of the peer review mechanism used within the OECD. The rationale for this particular case selection lies in the recognition that, on the one hand, soft law policy coordination within the EU shares a lot in common with the multilateral surveillance of the OECD⁸⁷, while on the other hand, valuable lessons for enhancing the effectiveness of EU evaluation mechanisms may be learned from the marked contrasts between the two frameworks.

*OECD peer reviews: functioning and rationale*⁸⁸

⁸⁷ Mutual surveillance rests on peer review. The similarity of the OMC to practices at the OECD and IMF has been noted (See Borrás, S. and Jacobsson, K., *The Open Method of Coordination and the New Governance Patterns in the EU*, *Journal of European Public Policy*, 2004, p. 188). Moreover, Wallace calls the OMC the ‘OECD technique’ (See Schäfer, A., *A New Form of Governance? Comparing the Open Method of Coordination to Multilateral Surveillance by the IMF and the OECD*, MPIfG Working Paper, 2005, p. 4.)

⁸⁸ For this evaluation, three of the OECD’s most well-known peer review processes are taken into consideration: the OECD Economic Surveys, the Environmental Performance Reviews and the Peer Reviews of the Development Assistance Committee (DAC). The evaluation provided here draws upon Kornblum, OECD – Joint Group on Trade and Competition, Schäfer and Thygesen. See Kornblum, E., *A Comparison of Self-Evaluating State Reporting Systems*, *International Review of the Red Cross*, 1995; OECD – Joint Group on Trade and Competition, *Peer Review: Merits and Approaches in a Trade and Competition Context*, Paris, OECD, 2002; Schäfer, A., *Resolving Deadlock: Why International Organisations Introduce Soft Law*, *European Law Journal*, 2006; Thygesen, N., *Comparative Aspects of Peer Reviews: OECD, IMF and European Union*, ed. Tanaka, K., *Shaping Policy Reform and Peer Review in Southeast Asia, Integrating Economies and Diversity*, Paris, OECD Publishing, 2008.

Of all the existing international frameworks where publicly available reviews are conducted, the OECD has the longest history of doing so.⁸⁹ Already at the inception of the Organization in 1961, its Economic and Development Committee (EDRC) began to monitor national economic policies and publish annual Economic Surveys, subjecting the draft reports to a full-day discussion, in what was to be the first systematic peer review process of any international institution. The idea of this annual review procedure was that ‘each country submit[ted] its economic situation and policies to the examination of all its partners. Ample opportunity [was] thus provided for discussion of major problems, and each country [was] confronted with an informed view on the impact of its policies on its neighbours’.⁹⁰

This clearly suggests that the conduct of peer reviews was strongly motivated by the notion that, in a world of ever-increasing interdependence, it is important for countries to be aware of the potential spill-over effects of national policies. Indeed, the peer review process initially drew its inspiration from this motive. However, if, in the early period, the focus of the process was to analyze and discuss short-term country outlooks and the macro-economic challenges related to that, in the course of time more attention was dedicated to studying a country’s structural policies. As a corollary, the main inspiration for such multilateral surveillance exercises came to be derived less from the ‘interdependence motive’, and more on the idea that it may be interesting for countries to learn from the particular experiences of other countries. The point of an examination of one State’s performance in a particular area by other States has increasingly been to help States improve their policymaking, adopt best practices and comply with established standards and principles.⁹¹

However, let us now move on from discussing the motivations for peer review towards the more specific issue of exactly how the peer review process within the framework of the OECD works.⁹² It should be noted that peer reviews are generated in a number of different domains, and that while there is no standardized peer review mechanism, all peer reviews share certain structural elements. Importantly, peer review is always a joint

⁸⁹ However, we even find an earlier version of such multilateral surveillance in the OECD’s predecessor, the Organisation for European Economic Cooperation (OEEC). In the framework of the European Payments Union (EPU), arguably the OEEC’s most successful part, a number of steps were established that to this day constitute the essence of multilateral surveillance: (1) missions by the organization to member countries aimed at evaluating the national economic situation and the government’s actions; (2) the issuing of policy recommendations for the improvement of the national economic situation; and (3) the hearing of government representatives and of the national central bank. See Schäfer, A., *Resolving Deadlock: Why International Organisations Introduce Soft Law*, *European Law Journal*, 2006, p. 199.

⁹⁰ Burgess, W., Cappier, B., Gore-Both, P. and Zolotas, X., *A Remodelled Economic Organisation*. A Report by the Group of Four, Established by the Resolution of January 14, 1960 of the Twenty Governments, Members or Associate Members, of O.E.E.C. and the Commission of the European Economic Community, Paris: Organisation for European Economic Cooperation, 1960, p. 20.

⁹¹ Thygesen, N., *Comparative Aspects of Peer Reviews: OECD, IMF and European Union*, ed. Tanaka, K., *Shaping Policy Reform and Peer Review in Southeast Asia. Integrating Economies amid Diversity*, Paris: OECD Publishing, 2008, p. 137 – 138.

⁹² This overview is heavily based on OECD, *Peer Review at a Glance*. Available on http://www.oecd.org/document/26/0,3343,en_21571361_37949547_38012314_1_1_1_1,00.html and on Schäfer, A., *A New Form of Governance? Comparing the Open Method of Coordination to Multilateral Surveillance by the IMF and the OECD*, MPIfG Working Paper, 2005.

operation involving the reviewed country, the examining countries and staff from the OECD Secretariat. The review itself is carried out by the committee, working party, or other body that has decided to undertake it, and officials in the relevant policy field from other countries are involved in the evaluation process. The individuals representing the reviewed country may include civil servants from Ministries and agencies, as well as delegates from different levels of government, in some cases Ministers. For the reviewed country, participation implies the duty to co-operate with the examiners and the Secretariat by, among other things, making documents and data available, responding to questions and requests for self-assessment, and hosting on-site visits.

Generally, a few countries are chosen as lead examiners for a particular review, while the rest of the group actively participates in the final discussion. The choice of lead examiners is usually based on a system of rotation among Member States, although particular knowledge of the country relevant to the review may be taken into account. The examiners represent the collective body carrying out the review and provide guidance in the collective debate. Their work includes examining documentation, taking part in discussions with the reviewed country and the Secretariat, and taking a lead speaker role in the debate in the collective body. The examiners may also participate in missions to the country. Lead examiners have a duty to be objective and fair, and free from any influence of national interest that would undermine the credibility of the peer review mechanism. Increasingly, civil society, business and labour are also invited to contribute to reviews.

The OECD Secretariat supports the process by producing documentation and analysis, organizing meetings and missions, stimulating discussion and maintaining continuity. The independence, transparency, accuracy and analytical quality of the Secretariat's work are essential to the effectiveness of the process. How the work is divided up between the Secretariat and the lead examiners, and the degree of interaction between them, varies widely. However, as a general rule, the Secretariat carries out the most labour-intensive part of the job, particularly if it has the most expertise in the topic under review.

Whatever the topic, peer review exercises are generally carried out on a regular basis, and result in a report that assesses accomplishments, spells out shortfalls and makes recommendations. This report is circulated to all OECD Member countries; this is where the bilateral preparatory phase ends. In the finalization of its reports (and this is where we move from the preparatory to the discussion phase), the OECD Secretariat engages in a full day of intensive, informal and non-adversarial debate amongst high-ranking officials from the reviewed country, representatives from other visiting national authorities and the relevant committee that commissioned the review. Although diplomatic in tone, the participants do not shy away from voicing concerns and criticisms.

These multilateral discussions are directed at redrafting the survey in a way that all governments can eventually agree to it. Both the influence and soft persuasion exercised by the countries that are present during this informal dialogue, as well as the recommendations that follow from it, are important for the effectiveness of the peer

review process, given the absence of sanctions or other legally binding enforcement mechanisms.

This one-day debate is followed by another one-day meeting between the OECD Secretariat and representatives of the reviewed country, so as to prepare a revised draft for final approval by the committee. The strongly comparative nature of such reports, with extensive use of benchmarking and, if possible, qualitative and quantitative assessments of performance (through the ranking of countries according to their performance and/or by ‘naming-and-shaming’), makes it difficult for a country to avoid recommendations based on relatively poor performance or relatively weak efforts to improve policies in a certain area. The final report culminating from this process is made available to the public.

Lessons from the OECD peer review mechanism

Evaluations regarding the effectiveness of the OECD peer review process for fostering mutual learning, furthering cooperation and/or encouraging compliance with guidelines mostly suggest that it is relatively successful, compared to similar work performed both by other international organizations, as well as by the EU. The particular usefulness of the OECD peer review process derives from several factors, which are identified in the following section, together with factors that may hamper the effectiveness of the peer reviews. The section will close by providing some essential general preconditions for peer reviews to be effective.

*Merits...*⁹³

An examination of the functioning of the OECD peer review process suggests that the merits of this particular working method are manifold.

1. The prescriptive use of *policy recommendations* in the peer review process has two advantages. The first is that this practice endows a process with a strong forward-looking orientation, as the focus is placed on what authorities can constructively do to address policy challenges and/or improve compliance with certain rules or guidelines. The second advantage is that any progress made may be conveniently monitored against such recommendations, although the precision with which this can be done of course partly depends on the quantitative/qualitative and vague/precise nature of such recommendations.

⁹³ This section draws upon Kornblum, OECD – Joint Group on Trade and Competition (2002) and Thygesen (2008). Kornblum, E., A Comparison of Self-Evaluating State Reporting Systems, International Review of the Red Cross, 1995; OECD – Joint Group on Trade and Competition, Peer Review: Merits and Approaches in a Trade and Competition Context, Paris, OECD, 2002; Thygesen, N., Comparative Aspects of Peer Reviews: OECD, IMF and European Union, ed. Tanaka, K., Shaping Policy Reform and Peer Review in Southeast Asia, Integrating Economies and Diversity, Paris, OECD Publishing, 2008.

2. The fact that OECD reviews expressly *follow up* on recommendations made by the relevant OECD committee in previous years should be considered a valuable asset: follow up reports provide valuable feedback for the review process more generally and, moreover, they contribute to ensuring that lessons learned during the review process are translated into concrete improvements.
3. The *regular basis* on which reviews are conducted pressures Member countries to continue paying attention to the issue(s) discussed and to actually and effectively act upon recommendations, in order to avoid another ranking at the bottom of the list, as well as to prevent the ‘shaming effect’ that goes with this. The regularity of peer reviews may thus motivate or accelerate improvements in domestic policymaking.
4. Attendance at reviews of *high-level, multi-institutional and permanent delegations* has two specific advantages. First, it ensures that the issue under review receives the attention of those who are in relatively powerful positions, which may contribute to enhancing the salience of the issue and, following from this, the probability of governmental action. Secondly, ongoing discussions within a loose-knit community of national regulatory officials forge close contacts between them and furnish ample opportunities for cognitive convergence on debated policy issues.
5. The *public availability of reports* may put additional pressure on the government to act, by stimulating public debate on the government’s performance regarding a certain issue(s). In addition, it mitigates the influence of vested interests by virtue of the transparency that is provided.
6. The *intensive, non-confrontational and non-legalistic nature of the interaction*, both between the relevant OECD committee with the reviewed country, as well as amongst member countries themselves, ensures that the peer review process is an ongoing and informal information-sharing experience, from which all partners may benefit. On the one hand, the non-adversarial nature of the process makes it easier for the reviewed country to accept criticism from peer countries; on the other hand, the continuous and intensive debates on a certain policy issue provide all countries involved – but particularly the lead examiners – with a continuous source of learning experiences and capacity-building.
7. The fact that *participation is voluntary* may be beneficial, in the sense that the participating country may be more willing to cooperate during the process. This may be reflected in an increased willingness on the part of the country to provide information and respond to questionnaires, and in its preparedness to openly comment on problems that are experienced and to ask peer countries for advice.

...and weaknesses

1. However, the fact that *participation is voluntary* also has its downsides. The very need to bring countries on board first of all implies that messages in reviews may be diluted, and some issues fudged, to ‘appease’ countries and ensure their continued participation and cooperation in such reviews. Moreover, it also results in the fact that recommendations may not be very far-reaching, so as to increase the probability of a country acting upon them.
2. We may question the usefulness of the *self-reporting aspect* within the OECD method. Although we may depart from the assumption that, given their voluntary participation, countries may be inclined to be open and complete in their reports, as suggested previously, one may argue that some caution is warranted, since countries may have several motivations for refraining to report truthfully.
3. It is believed that OECD reviews are often *overly standardized*. However, as we have seen in the preceding chapter of this report, countries may differ greatly regarding their institutional set-up, administrative structures and political cultures, which means that a specific approach may perhaps fit one or several countries, but certainly not all. An approach that does not pay sufficient attention to the differences in domestic situations cannot be expected to produce optimal results.
4. It has to be recognized that the OECD peer review process functions at *high cost*. Although this is not so much inherent to the evaluation mechanism *per se*, it is not always self-evident that the benefits gained (if at all) do indeed offset the costs involved. Efficiency requirements may force us to conduct at least some type of cost-benefit analysis, before actually conducting a peer review study, the effectiveness of which can never (fully) be determined from the outset.

*Preconditions for peer review to be effective*⁹⁴

As a conclusion to this section, let us now move on to some essential preconditions that, as suggested by the experience of the OECD peer review process, must be present for peer review exercises to be effective.

1. *Participating countries must share the same views* on the standards or criteria against which to evaluate performance. A strong common understanding on these elements is necessary to prevent uncertainty or backtracking during the process.
2. A peer review can function properly only if there is an *adequate level of commitment by the participating countries*. This includes not only the notion that countries should supply enough resources to carry out the peer review process, but also that States should be fully engaged in the process, both when acting as reviewers, as well as when being reviewed.

⁹⁴ The preconditions provided are identified at the website of the OECD, [Peer Review. Why Does It Work? Available on \[http://www.oecd.org/document/27/0,3343,en_21571361_37949547_37958363_1_1_1_1,00.html\]\(http://www.oecd.org/document/27/0,3343,en_21571361_37949547_37958363_1_1_1_1,00.html\)](http://www.oecd.org/document/27/0,3343,en_21571361_37949547_37958363_1_1_1_1,00.html).

3. Since peer review is, by nature, a co-operative and non-adversarial process, *mutual trust* is important for its success. While the peer review process itself can contribute to confidence building, a large degree of trust and value sharing among the participants should be present from the beginning, in order to facilitate the disclosure of data, information and documentation that are essential to the process.
4. The *credibility of the peer review process* is essential to its capacity to influence countries and, hence, to its effectiveness. The OECD's independent Secretariat, designated examining countries and a multilateral committee process have each been identified as helping to ensure this credibility. Moreover, the involvement of the reviewed country in the process, and its ownership of the outcome are the best guarantees that it will ultimately endorse the final report and implement its recommendations. However, this involvement should, of course, not go so far as to endanger the fairness and the objectivity of the review.

Obviously, an awareness of both the merits and weaknesses associated with the OECD peer review process, as well as the preconditions necessary for peer review to be effective, is imperative when re-thinking the functioning and design of similar evaluation mechanisms used in the EU. We now leave the OECD context behind for the time being, and move on to examining the evaluation mechanisms employed in the first pillar of the EU.

3.2 Monitoring and evaluation mechanisms in the first pillar of the EU

The types of monitoring and evaluation mechanisms employed in the EU are, to a greater or lesser extent, similar to those used in the OECD. However, we have to acknowledge the fact that, rather than being conducted as an activity *per se*, peer review in the EU is part of a much wider governance framework, and that such 'soft' multilateral surveillance mechanisms exist together with and/or instead of mechanisms that feature some more or less 'hard' elements. Before moving on to an examination of the functioning and effectiveness of some specific evaluation mechanisms in the next chapter, the remainder of this chapter sets the scene by providing a general overview of evaluation mechanisms used within the EU's first pillar. We pay specific attention to the increasingly-used peer review mechanism that takes on a particularly important role within the Open Method of Coordination (OMC).

'Hard' and 'soft' methods of evaluation

Within the context of first pillar cooperation, monitoring and evaluation have become common policy tools. Depending on the more or less legalistic nature of the instrument that is evaluated, the Member States have accepted the use of quite far-reaching evaluation mechanisms, aimed at enforcing compliance with binding rules or promoting convergence towards non-binding guidelines. It is self-evident that the use of coercive mechanisms, where hard law is concerned (see the discussion in the previous chapter),

cannot be extended to those policy domains where there has merely been a ‘partial delegation of powers’ to the EU level.

In those areas where the Member States remain reluctant to shift legislative competences to the supranational level, but where a shared incentive to act has nonetheless resulted in the development of some sort of ‘soft law policy cooperation’, monitoring and evaluation mechanisms take on an appearance that may be far removed from the coercive nature of mechanisms such as the European Commission and the ECJ acting as guardians of the Treaties. Here, compliance with non-binding guidelines (and consequently convergence towards collective goals) may be promoted through peer review exercises, in which benchmarking, the sharing of ‘best practices’, multilateral surveillance and peer pressure constitute essential elements. The mutual learning process that is desired may be backed up by the potentiality of recommendations being issued to Member States whose performance is deemed unsatisfactory, and/or by the use of ‘soft incentives’ aimed at stimulated compliance, such as ‘naming and shaming’, or media publicity.

The Open Method of Coordination: peer review at the core

Although, over time, peer review activities emerged as an essential aspect of an increasing number of domains that are subject to policy coordination between the Member States, their significance as a policy tool was perhaps most clearly recognized by the Lisbon European Council, which formally established the Open Method of Coordination (OMC) as a ‘new’ method of governance. The Lisbon conclusions presented the OMC as a means of achieving greater convergence in politically sensitive areas, where ideological differences between the Member States rendered harmonization impossible, and as a means of disseminating ‘best practices’ about appropriate public policies, identified through a process of mutual learning.⁹⁵

Before examining more closely the effectiveness of multilateral surveillance within the OMC, let us first consider the essential components of this method of coordination. In its ‘ideal-typical’ form, the OMC is repeated regularly and proceeds as follows: the Council of Ministers agrees on common objectives (with indicators and benchmarks, ‘where appropriate’), which are drafted primarily by the European Commission. Subsequently, the Member States, using the agreed list of indicators and benchmarks, translate the guidelines into domestic policies, and report on both the policies they have implemented and planned, as well as the progress made. The Commission then assesses and compares their efforts, identifies best practices and formulates recommendations for each Member State.

This evaluation process is accompanied by peer review exercises involving the Member States, as well as regional, municipal and non-governmental actors. The peer review process aims to provide learning opportunities that feed back into the development of national policy and the re-formulation of guidelines. The results of the evaluation process

⁹⁵ Radaelli, C., *The Open Method of Coordination: A New Governance Architecture for the European Union?*, SIEPS Reports, Nr 1, 2003, p. 20.

are laid out in a joint report that must be approved by the Council. In the follow up stage, the Member States are expected to implement the measures recommended to them.⁹⁶

It is important to note that this ideal-type of OMC allows for a considerable diversity of formats, rationales and results across policy areas: while some domains have been characterized by a deliberate attempt to use the OMC as the main working method (for example, the Broad Economic Policy Guidelines (BEPGs), the European Employment Strategy (EES) and Social Inclusion), other areas are only vaguely similar to the ideal-typical model (for example, innovation, information society and health care).

If this diversity within the OMC ‘template’ across policy domains is already very clear from an examination of the range of guidelines and indicators established in different fields, it appears that peer review also means different things in different policy areas. In some cases (for example, innovation), peer review has thus far not been extensively utilised, whereas in other cases (for example, employment policy), comprehensive peer review exercises have become institutionalized and the emphasis on peer pressure as a means of improving national policy making has been significant. The use of recommendations as an additional way of putting pressure on Member States is not generally accepted; they are applied only in the context of the BEPGs and in the EES.⁹⁷

Diversity is also visible where benchmarking is involved. Although its practice is widely utilised, there are several different political contexts in which this is done. While in some areas the diversity of opinions of what constitutes ‘best practice’ may be such that Member States merely manage to agree on what is ‘worst practice’ (for example, in the field of taxation), in other domains, differences between Member States may be less pronounced, and policymakers may be more ready to engage in learning from ‘best practices’ (for example, in the domain of employment policy).⁹⁸

Finally, the involvement of social partners, local actors, civil society representatives and national parliaments within the evaluation process varies across policy domains. While there is no participation at all in some policy domains (for example, in relation to the

⁹⁶ Kerber, W. and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 2007, p. 231; Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 14-15.

⁹⁷ Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 34 – 36; Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008, p. 508; Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, *La Follette School Working Paper Series*, No. 2007-031, 2007, p. 4.

⁹⁸ Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 34 – 36. ; Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008, p. 508; Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, *La Follette School Working Paper Series*, No. 2007-031, 2007, p. 4.

BEPGs), we find greater (but, in comparison with the ideal-type OMC, still limited) participation in other policy areas (for example, where social inclusion is concerned).⁹⁹

Assessing the effectiveness of the OMC

How effective have these different OMC processes been for achieving the envisaged objectives? As mentioned, the OMC was hailed as an institutional arrangement designed to achieve greater convergence of national policies, while respecting national diversity, as well as enabling the Member States to learn more about policy. As it is beyond the scope of this report to provide a thorough assessment of the extent to which the OMC *per se* may be considered effective, our focus will be on identifying some theoretical and empirical aspects related to the effectiveness of peer review for fostering policy learning and promoting policy convergence.

Theoretical promises...

Different schools of thinking in academic literature hold that, in theory, the potential for the OMC to bring about policy learning should not be underestimated. Two different types of policy learning are identified.¹⁰⁰

1. Top-down policy learning:
 - a. *Shaming* has the potential to bring about learning: as Member States seek to avoid negative criticism in peer reviews, they will feel compelled to implement national policies aimed at achieving compliance with guidelines and/or recommendations.
 - b. *Diffusion* stimulates learning: either through mimesis (according to which Member States copy ‘best practices’ identified through peer review), or through discursive transformation (according to which peer review processes transform national discourses and thus national policy).
 - c. The *creation of policy networks* through the presence of government officials and non-governmental stakeholders at frequent peer review exercises is expected to lead to the diffusion of new ideas and the emergence of a common set of policy positions.

⁹⁹ Kerber, W. and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 2007, p. 236; Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 37-38.

¹⁰⁰ Trubek, D. and Trubek, L., Hard and Soft Law in the Construction of Social Europe: The Role of the Open Method of Coordination, *European Law Journal*, 2005, p. 356 -359; Kerber, W. and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 2007, p. 229.

- d. The *peer pressure* inherent in any such peer review exercise is considered a potentially powerful tool for encouraging Member States to comply with guidelines and/or recommendations.

2. Bottom-up policy learning:

- a. Peer review processes foster *experimentation, deliberation and learning*. The diversity of the EU is a valuable asset, as many different policies will be utilised simultaneously. The sharing of information on these different policies, the comparison of different national experiences and the identification of ‘best practices’ during peer review processes, are expected to be useful tools to help Member States find new or better solutions to certain policy problems and, thus, to help them converge towards common goals.

... and the empirical reality

From a theoretical point of view, the potential of the OMC to bring about policy learning and policy change is significant. However, the empirical research on the effectiveness of the OMC shows a rather sceptical picture about the way in which this potential has been realized. There seems to be a general perception that, although OMC processes may have contributed to some convergence of perceptions, orientations, interpretative schemes and the defining of problems, there seems to be little evidence for first order policy learning or direct policy transfer at the national level and, thus, for actual policy convergence.¹⁰¹

1. A first problem is that, in most OMC processes, *the participation of social partners, local actors, civil society representatives or even national parliaments has been weak or non-existent*. Peer reviews have been described as a learning process for a limited community of experts, and thus described as exercises in ‘expertocratic deliberation’, rather than open processes involving the active engagement of the main parties affected by the issue at stake. As stakeholder buy-

¹⁰¹ López-Santana, M., The Domestic Implications of European Soft Law: Framing and Transmitting Change in Employment Policy, *Journal of European Public Policy*, 2006, p. 494; Zeitlin, J., The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels: Peter Lang, 2005, p. 457, 472; Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008. We should note that it is difficult to assess the effectiveness of the OMC in terms of its legislative output or outcomes. First of all, it is hard to establish what would have happened without it and, secondly, its possible effects only become visible in the medium or long term (Schäfer, A., A New Form of Governance? Comparing the Open Method of Coordination to Multilateral Surveillance by the IMF and the OECD, MPIfG Working Paper 04/05, 2005). Even the official evaluation of the EES after five years stresses the difficulty of isolating its effects from the general economic cycle on the one hand, and from policy initiatives that originated from different sources on the other (Commission of the European Communities (CEC), Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. Taking Stock of Five Years of the European Employment Strategy, COM (2002) 416 final, 17 July 2002). In other words, it is difficult to establish the causal relationship between the OMC and perceived policy changes.

in is both key to the effectiveness of policy learning, as well as important for stimulating compliance through the ‘watchdog function’, participation thus falls short of the ideal-type of participatory governance that had been designed at Lisbon.¹⁰²

2. The *political nature of mutual learning processes* constitutes an additional hindrance to effective policy learning. The agreements to use a certain range of indicators, to regard a certain policy as ‘best practice’, and to issue recommendations, are decisions that result from complex political bargaining processes. The engagement of national policymakers in peer reviews may be motivated not so much by their desire to genuinely learn from each other, but rather is conditioned by considerations related to power. In short, peer review exercises may not be as apolitical as we might expect, and political conflicts and power struggles may seriously hinder policy learning processes.¹⁰³
3. The method of benchmarking and searching for best practices implicitly assumes that a ‘best practice’ for all Member States can be identified. However, *domestic institutional structures and political cultures differ throughout the Union* and the probability of policy transfer to occur is therefore limited. Until now, awareness that learning involves a ‘context-sensitive ‘lesson-drawing’ approach’, according to which contextualized learning takes the specific institutional conditions of Member States into consideration, has not been sufficiently present. This constitutes a clear hindrance to policy learning and explains the only limited evidence of successful instances of policy transfer.¹⁰⁴

¹⁰² Casey, B. and Gold, M., Peer Review of Labour Market Programmes in the European Union: What Can Countries Really Learn from One Another?, *Journal of European Public Policy*, 2005, p. 37; Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008, p. 502-503; Kerber, W. and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 14, 2007, p. 236; Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 40; Zeitlin, J., The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels: Peter Lang, 2005, p. 476.

¹⁰³ Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008, p. 504; Kerber, W., and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 14, 2007, p. 235 - 236; Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 40; Zeitlin, J., The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 474.

¹⁰⁴ Heidenreich, M. and Bischoff, G., The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?, *Journal of Common Market Studies*, 2008, p. 520; Kerber, W. and Eckardt, M., Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism, *Journal of European Public Policy*, 14, 2007, p. 234; Radaelli, C., The Open Method of Coordination: A New Governance Architecture for the European Union?, *SIEPS Reports*, Nr 1, 2003, p. 42.

4. There seems to be a generally shared perception that the limitations of mutual learning also stem from some *very practical defects related to the OMC procedures and instruments per se*. To mention just a few, full agendas, a very tight timetable for peer review, voluntary (and thus limited) participation of the Member States in peer review exercises, and fragmentary coverage of practices for examination, have each been identified as factors hindering mutual learning. The persistent lack of comparable data, due partly to a deliberate attempt by Council representatives to limit the comparability and comparison of national performances, and the questionable relevance or lack of clear causal significance of the indicators established, should also be regarded as factors decreasing the effectiveness of the peer review mechanism.¹⁰⁵
5. Finally, some authors find that *the incentives for the Member States to implement guidelines and/or policy recommendations are unclear and are often weak*. It is argued that the OMC does not constitute a strong enough instrument to effectively stimulate compliance by the Member States.¹⁰⁶ Not all scholars agree that the solution to overcome this problem is to strengthen coercive sanction mechanisms as a way of enhancing compliance.¹⁰⁷ Others argue that the effectiveness of these tools, and particularly the ‘naming-and-shaming’ approach, depends on the perceived legitimacy of the recommendations, national sensitivity to criticism from the EU, and the domestic visibility (for example, through media coverage) of the process.

3.3 Existing evaluation mechanisms in the third pillar on criminal law

Despite the absence of infringement proceedings in the third pillar, several mechanisms of evaluation have been established over the last 15 years.¹⁰⁸ We have identified the following existing evaluation mechanisms:

1. Evaluation by the Commission of the implementation of legislation by the Member States;
2. Peer evaluation, on the basis of the Joint Action 5 December 1997;

¹⁰⁵ Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 474 - 475.

¹⁰⁶ Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 524-525; Kerber, W. and Eckardt, M., *Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism*, *Journal of European Public Policy*, 14, 2007, p. 236.

¹⁰⁷ López-Santana, M., *The Domestic Implications of European Soft Law: Framing and Transmitting Change in Employment Policy*, *Journal of European Public Policy*, 2006; Van Vliet, O. and Koster, F., *Europeanisation and the Political Economy of Active Labour Market Policies*, Paper prepared for the NIG Annual Work Conference 2008, Enschede, 20-21 November, 2008.

¹⁰⁸ Communication from the Commission to the Council and the European Parliament, *Evaluation of EU Policies on Freedom, Security and Justice*, COM (2006) 332 final, Brussels, 28.6.2006, Annex 2, *Current practice for monitoring and evaluating EU policies on freedom, security and justice*. Note that the Commission includes another categorisation: evaluation of programmes, of legislation and of policies.

3. The Schengen evaluation mechanism;
4. Evaluations undertaken in preparation for EU enlargement;
5. Follow up of judgments of the ECJ.

Evaluation by the Commission of the implementation of legislation by the Member States

These evaluations find their legal bases in the specific legal instruments.¹⁰⁹ These contain a provision that merely obliges the Commission to draw up a report on the transposing of legislation of the Member States submitted to the Commission.¹¹⁰ The legal instruments do not give any indication as to either how to evaluate, or how to prepare for, the drafting the report. As an example, Article 14, paragraphs 1 and 2 of Framework Decision 2003/577 provides:

‘1. Member States shall take the necessary measures to comply with the provisions of this Framework Decision before 2 August 2005.

2. By the same date Member States shall transmit to the General Secretariat of the Council and to the Commission the text of the provisions transposing into their national law the obligations imposed on them under this Framework Decision. On the basis of a report established using this information and a written report by the Commission, the Council shall, before 2 August 2006, assess the extent to which Member States have complied with the provisions of this Framework Decision.’

¹⁰⁹ Report from the Commission based on Article 11 of the Council's framework Decision of 29 May 2000 on increasing protection by criminal penalties and other sanctions against counterfeiting in connection with the introduction of the euro {SEC(2001)1999} /* COM/2001/0771 final */; Report from the Commission based on Article 14 of the Council Framework Decision of 28 May 2001 combating fraud and counterfeiting of non-cash means of payment {SEC(2004) 532} /* COM/2004/0346 final */; Report from the Commission based on Article 6 of the Council Framework Decision of 26 June 2001 on money laundering, the identification, tracing, freezing, seizing and confiscation of instrumentalities and the proceeds of crime {SEC(2004)383} /* COM/2004/0230 final */; Report from the Commission - Implementation by Member States of the Convention on the Protection of the European Communities' financial interests and its protocols Article 10 of the Convention {SEC(2004) 1299} /* COM/2004/0709 final */; Report from the Commission based on article 11 of the Council Framework Decision of 13 June 2002 on combating terrorism {SEC(2004)688} /* COM/2004/0409 final */; Report from the Commission on the basis of Article 18 of the Council Framework Decision of 15 March 2001 on the standing of victims in criminal proceedings {SEC(2004)102} /* COM/2004/0054 final/2 */; Report from the Commission on the Legal Transposition of the Council Decision of 28 February 2002 Setting up Eurojust with a View to Reinforcing the Fight Against Serious Crime {SEC(2004)884} /* COM/2004/0457 final */; Report from the Commission on national measures taken to comply with the Council Framework Decision of 13 June 2002 on Joint Investigation Teams {SEC(2004) 1725} /* COM/2004/0858 final */; Report from the Commission based on Article 34 of the Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States {SEC(2005) 267} /* COM/2005/0063 final */.

¹¹⁰ One may question whether these types of provision are practical, or whether a common legal basis, such as provided in Article 70 TFEU, would not be better.

The Commission interprets these provisions as the basis for an evaluation of the implementation of the instruments. None of the legal instruments gives any indication about the standards and topics for evaluation, other than that it should relate to formal implementation into national legislation.

Whereas checking whether Member States have implemented legal instruments finds a direct legal basis in the relevant instruments, two other practices have also developed. The first is the ‘scoreboard’ mechanism, which gives an overview of compliance and implementation of the commitments in the multi-annual programmes of the third pillar.¹¹¹ It indicates the objectives and deadlines set and, in each case, the responsibilities assigned to launch, advance and complete the process.

A further mechanism is the creation of a Forum for discussing EU justice policies and practice, which was established last year.¹¹² The potential of the Justice Forum is promising. However, it is too early to evaluate its effectiveness.

Peer evaluation on the basis of Joint Action 97/827

The Joint Action of 1997 establishes a general mechanism of evaluation of the whole third pillar acquis, although it was formally limited to organised crime.¹¹³ The reports that result from this type of evaluation are conducted with a broader scope, focusing not only on implementing legislation, but also on the practical application of the instruments concerned. However, it is unclear how the topics for evaluation have been selected and according to which research plans they have subsequently been conducted.

Since 1998, peer evaluations on a mutual basis have taken place in four different rounds. The topics were decided by the Multidisciplinary Group on Organised Crime. The first round dealt with mutual assistance in criminal matters;¹¹⁴ the second with Law Enforcement and its Role in the Fight Against Drug Trafficking;¹¹⁵ the third focused on the exchange of information and intelligence between Europol and the Member States

¹¹¹ COM (2008) 373, Communication from the Commission to the Council and the European Parliament - Report on implementation of the Hague programme for 2007 {SEC (2008) 2048} {SEC(2008) 2049}.

¹¹² See Communication from the Commission on the creation of a Forum for discussing EU justice policies and practice /* COM/2008/0038 final.

¹¹³ Joint Action 97/827 of 5 December 1997 adopted by the Council on the basis of Article K.3 of the Treaty on European Union, establishing a mechanism for evaluating the application and implementation at national level of international undertakings in the fight against organized crime, OJ L 344, 15.12.1997, p. 7–9.

¹¹⁴ Council of the European Union, Final Report on the first evaluation exercise – mutual legal assistance in criminal matters, Doc. 8648/01, 10 May 2001. See also Joint Action 98/427 of 29 June 1998 on good practice in mutual legal assistance in criminal matters, OJ 1998, L 191/1.

¹¹⁵ Second round of Mutual Evaluations Law Enforcement and its Role in the Fight Against Drug Trafficking, Report on the second five evaluation visits (Italy, Greece, Denmark, Spain and Ireland) Brussels 10 December 2002, 12359/3/02.

and among the Member States; and the fourth dealt with the practical application of the European Arrest Warrant.¹¹⁶

The reports give the impression that the focus was predominantly on implementing legislation. The Final Report on the Third Round of mutual evaluations on ‘Exchange of information and intelligence between Europol and the Member States and among the Member States respectively’ merely states: ‘the primary purpose of the third round of mutual evaluation should be to evaluate the application and implementation at national level of instruments dealing with law enforcement of the resulting legislation and the practices at national and international cooperation. In particular, the evaluation was to assess cooperation and coordination between different law-enforcement structures and operational practices. Overall, the main focus of the evaluation was to be practical day-to-day cooperation between different units at both national and international level.’¹¹⁷

In the aftermath of 9/11, the Justice and Home Affairs Council decided to develop an easier and swifter form of evaluation mechanism than the earlier Joint Action 97/827 for the evaluation of the legal systems and their implementation at national level in the fight against terrorism. This resulted in Council Decision 2002/996 of 28 November 2002.¹¹⁸ In comparison to Joint Action 97/827, there is a greater obligatory requirement present (Article 1 Decision 2002/996), which is also demonstrated by the strict deadlines for each step. The coordination clearly lies with CATS (Article 2) in relation to the specific subject matter of the evaluation, the frequency of evaluations and the order in which Member States are to be evaluated. The Decision provides for an evaluation team composed of experts suggested by the Member States. A questionnaire is prepared by the Presidency. After receipt of the questionnaire the team visits the Member State. The report subsequently drafted by the team is discussed in CATS, with the comments of the Member States evaluated. Article 9 Decision 2002/996 stipulates that the report shall be ‘at least a restricted document.’

The Schengen evaluation mechanism

In order to facilitate and maintain mutual trust, the Schengen Member States set up a Standing Committee in 1998. Its mandate is set out in a decision of the Schengen Executive Committee (SCH/COM-ex (98) 26def) and consists of two separate tasks:

1. Verification of whether all preconditions for application of the Schengen acquis have been met by Member States wishing to join Schengen (‘putting into effect/application’).
2. Verification that the Schengen acquis is being correctly applied by the Member States implementing the acquis (implementation).

¹¹⁶ See, for example, Evaluation report on the fourth round of mutual evaluations, ‘the practical application of the European Arrest Warrant and corresponding surrender procedures between Member States’ report on Spain, Brussels, 6 June 2007, 5085/2/07, REV2.

¹¹⁷ Brussels, 23 November 2007, 13321/3/07, p. 3.

¹¹⁸ Council Decision 2002/996 of 28 November 2002 establishing a mechanism for evaluation of the legal systems and their implementation at national level in the fight against terrorism, OJ 2002, L 349/1.

Within the Schengen context, a distinction is therefore made between ‘putting into effect’ and ‘implementation’. The first verification task aims to determine whether the conditions for mutual trust are met, whilst the second aims to maintain mutual trust by checking for the correct implementation of the acquis. The evaluation within the Schengen context is performed by the Member States, with the Commission only participating as an observer.

Earlier this year, the Commission presented two identical proposals for a Decision and a Regulation on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis.¹¹⁹ As expressed by the Commission, the main objective of the proposal is to establish a legal framework for evaluating the correct application of the Schengen acquis. If the proposals were to be adopted, it will lead to a certain degree of institutionalisation of the evaluation and monitoring mechanisms. It builds upon the existing evaluation mechanisms and gives a more prominent role to the Commission, whereas under the current Schengen regime the Council is the responsible institution. The proposals provide for expert teams to undertake on-site visits, composed of experts from a List of Experts. The mechanisms will also make use of questionnaires and produce evaluation reports. The Commission expressly refrained from proposing unannounced on-site visits. A follow up to the findings and recommendations of the Evaluation report is provided.

Evaluations in preparation of enlargement

Another mechanism of a more general nature is the system set up in 1998 to evaluate applicant States in relation to their compliance with the acquis of the third pillar.¹²⁰ In this evaluation process, the relevant judicial systems are assessed, including criteria related to the rule of law. This type of evaluation is focused on the following indicators pointing to the reliability of the judicial system:

- independence of the judiciary;
- efficiency of justice (in particular, the organisation of the judicial system, management of cases and delays);
- quality of personnel (recruitment and training of judges and other judicial personnel, career development, transparency);

¹¹⁹ Decision on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 105 final, 2009/0032 (CNS) and Regulation on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 102 final, 2009/0032=3 (CNS). The double evaluation mechanism is due to the fact that part of the Schengen acquis belongs to the first pillar, and part to the third.

¹²⁰ Joint Action 98/429 of 29 June 1998 adopted by the Council on the basis of Article K.3 of the Treaty on European Union, establishing a mechanism for collective evaluation of the enactment, application and effective implementation by the applicant countries of the acquis of the European Union in the field of Justice and Home Affairs, OJ L 191, 7.7.1998, p. 8–9. See also Council Regulation 622/98 of 16 March 1998 on assistance to the applicant States in the framework of the pre-accession strategy, and in particular on the establishment of Accession Partnerships, OJ 1998, L 85/1.

- access to justice (for example, information systems with regard to individuals involved in criminal proceedings and legal aid);
- respect for underlying guarantees (particularly the fight against corruption).

Follow up of judgments of the ECJ

In the context of the first pillar, the Commission supervises whether a Member State found in violation of the Treaties by the ECJ has implemented the findings of the Court. If the Commission is not satisfied, it may bring a case against that Member State. Where the ECJ holds that a Member State is still found to be in violation, despite a previous judgment of the Court, it may impose a financial penalty..

In the context of the third pillar, such a two-pronged infringement procedure does not exist. In the third pillar, there have been two cases in which the Court has nullified, on the application of the Commission, a Framework Decision (Commission v. Council, C-176/03 and C-440/05), on the basis of Article 35, paragraphs 1 and 6 TEU, by which Framework Decisions 2003/80 and 2005/667 were nullified. The nullification takes immediate effect. A follow up need not be monitored. Similarly, the Court nullified several Common Positions on combating terrorism.

The other category of Court decisions follow after a preliminary reference by a national court, on the basis of Article 35, paragraphs 1 TEU. These preliminary rulings take place under the aegis of what the Court refers to as, 'judicial cooperation' between the Court and national courts. The follow up to the ruling of the Court must be safeguarded by the national court. There is no role provided for the Commission. It may not bring a case against a Member State alleging that that State has not complied with a preliminary ruling. The Court cannot impose fines. On a more practical level, the Court does request the national court, when transmitting its judgment, to inform the ECJ of any follow up to the Court's decision. This information is collected by the Court and disseminated in the so-called 'suivi des affaires préjudicielles.'

One will notice that the third theoretical possibility, that of bringing a dispute on the interpretation of Union law before the Court on the basis of Article 35, paragraph 7 TEU, has never been used.

4. Monitoring and evaluation mechanisms at work in the first pillar

Having set the scene regarding both compliance in general, as well as the use of evaluation mechanisms within the EU's first and third pillars from a more abstract perspective, we now move on to an evidence-based analysis of such mechanisms within some specific areas.

The focus of this chapter will be on the functioning of the evaluation mechanism within the European Employment Strategy (EES). The selection of this particular area is due to the fact that it is exemplary in several respects, and that its analysis – even if focused on the specific functioning of the Mutual Learning Programme (MLP) that constitutes the central tenet of evaluation in this policy domain – leads to valuable insights into the functioning and effectiveness of various monitoring and evaluation mechanisms within different contexts.

The analysis on the Mutual Learning Programme constitutes the basis for our recommendations, which are presented later in this chapter, on how to improve the functioning of evaluation mechanisms. In order to demonstrate the wider validity of these recommendations, we have noted how factors identified either as contributing to or hindering, the optimal functioning of the MLP, often also figure in discussions regarding the effectiveness of evaluation mechanisms in other policy areas. We will concentrate particularly on the diverse instruments for evaluation employed within the domain of anti-discrimination policy, where Member States are required to ensure the correct implementation of the Racial and Employment Equality Directives.

To further substantiate our conclusions and recommendations, attention is also given to the evaluation mechanisms used to ensure compliance with the fiscal rules within the Stability and Growth Pact (SGP), and the way in which the application by the Member States of the Judgment Regulation was evaluated.

For a proper analysis of these case studies, as well as for the analysis of the third pillar practices in the next chapter, the following questions are considered relevant:

Regarding the *evaluated instrument(s) per se*:

- How can the nature of this instrument be described? Are we dealing with a rule-based framework, or with a context characterized by soft law policy coordination? In other words, what is it exactly that is subject to evaluation?

Regarding the *functioning* of the monitoring and/or evaluation mechanism(s) applied within this framework:

- Who are the (public and private) actors involved in the monitoring and evaluation exercises, and what are their respective roles?
- What methodology is used for monitoring and evaluation?
- What is the frequency of the monitoring and evaluation exercises?

- How are the results of the evaluation circulated? To what extent is the information publicly accessible?
- Does (do) the evaluation mechanism(s) provide for recommendations to be made as a result of the findings of the evaluation exercise? If yes, is compliance with these recommendations by the respective Member States followed up?
- Are sanctions provided for in case of non-compliance?

Regarding the *effectiveness* of the respective monitoring and/or evaluation mechanisms:

- Is there a link between the level of compliance with rules or guidelines and (if applicable) recommendations on the one hand, and the way in which the evaluation was carried out on the other? What factors could be regarded as stimulating or hindering compliance with the outcome of the evaluation mechanism?

4.1 Characterizing the European Employment Strategy

As noted above, the EES constitutes the most developed and prominent example of the OMC. Launched by the Luxembourg ‘Jobs’ Summit in 1997, the EES, or ‘Luxembourg Process’, was a political compromise designed to implement the provisions of the Treaty of Amsterdam, which authorized the EU to coordinate the policies of Member States towards the achievement of a ‘high level of employment’ as a matter of common concern, while not granting it any new legislative competences or spending powers. The EES sought to meet this objective by providing for ‘a new iterative process of benchmarking national progress towards common European objectives, supported by organized mutual learning’.¹²¹

The process, as it developed during the early years of the EES, consisted of five main steps in an annual cycle:¹²²

1. The adoption by the Council of common but formally non-binding European Employment Guidelines (EEGs), targets and indicators, based on intra- and extra-European benchmarking;
2. National Action Plans for Employment (NAPs/empl), through which Member States assess their relative progress towards the common objectives and targets, and propose corrective action in areas of shortfall;
3. Peer review of the NAPs/empl by Member State representatives in the Employment Committee (EMCO), supported by a contextualised exchange of good practices at the national level;

¹²¹ Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, La Follette School Working Paper Series, No. 2007-031, 2007, p. 2.

¹²² Idem.

4. A Joint Employment Report and a set of country-specific recommendations, proposed by the Commission, reviewed by EMCO, and approved by the Council;
5. Review and revision of the guidelines, targets, indicators and procedures.

Analyzing the characteristics of this process in terms of legalization (that is, assessing the process in terms of obligation, precision and delegation), Radulova suggests that the EES must be considered a relatively¹²³ ‘heavy’ OMC process.¹²⁴ This qualification derives, first of all, from the fact that, rather than being the product of Council decisions, the process is ‘treaty-based’: as noted above, the very obligation for Member States to coordinate their employment strategies was expressly agreed to by the Amsterdam European Council and integrated into the Treaty Establishing the European Community (TEC).¹²⁵ Both this notion and the idea that this type of obligation is absent in most other OMC processes,¹²⁶ give rise to the conclusion that the level of obligation inherent in the EES is relatively high.

Secondly, it is suggested that relative to other OMC processes, the EES demonstrates a high degree of precision. As both the long-term objectives, as well as the means to achieve these objectives, are specified in the Employment Guidelines, the leeway for interpretation and discretion allowed to the Member States is low in comparison with other OMCs. However, a *sec* examination of the Employment Guidelines (both past and present) demonstrates that not all guidelines are equally precise in their definition of targets and envisaged action: for example, whereas some EEGs spell out quantitative targets and specific strategies for action to be undertaken by the Member States, others merely require the Member States to take ‘appropriate’ steps to advance towards a what is already a quite vaguely defined goal.

Thirdly, the EES operates through a relatively high degree of delegation. There is a considerable degree of managerial delegation from the national level to the supranational level, visible in the fact that the Member States have entrusted the European Commission with significant authority to deal with the procedural processes of the EES. Given its tasks of issuing guidance notes, monitoring the process of implementation and collecting and interpreting data about performance by the Member States, the Commission clearly acts as a mediator and the operational interface for the EES.

¹²³ The concept of ‘relatively’ refers to the comparison of the EES OMC with other OMC processes, rather than with policymaking processes in general, many of which may, of course, be characterized by the making of ‘hard’ law.

¹²⁴ Radulova, E., *Governing Childcare policy through the Open Method of Coordination: the career of a policy solution in the Netherlands and at the EU level*. PhD thesis, Maastricht University, forthcoming (2010), chapter 3.

¹²⁵ See Articles 125-130 TEC. The obligation for Member States to coordinate their employment policies is made explicit in Article 125 TEC, which provides that ‘Member States and the Community *shall* . . . work towards developing a coordinated strategy for employment’, and in Article 126 TEC, which specifies that ‘Member States . . . *shall* regard promoting employment as a matter of common concern and *shall* coordinate their action in this respect with the Council’ (italics added). Article 128 TEC lays down the specific steps that are part of the iterative process described.

¹²⁶ A notable exception to this is the coordination of economic policies through the Broad Economic Policy Guidelines (BEPGs) under Articles 98 and 99 TEC.

Moreover, the Commission's role of proposing recommendations to Member States whose performance is deemed unsatisfactory also gives the process a relatively high degree of enforcement delegation, although these recommendations must, of course, be adopted by the Member States in the Council.

However, although unlike most OMC processes,¹²⁷ the EES contains some elements that we may associate with 'hard law', it essentially remains a 'soft law policy coordination mechanism'. The Employment Guidelines are non-binding, and sanctions for those Member States that fail to meet targets or follow up recommendations are non-existent. As a matter of fact, the only way in which the Member States may be moved to comply with guidelines or recommendations is through 'peer pressure'. As we will see in the following section, the peer review procedure inherent in the design of the EES was consequently hailed as an important and promising tool for contributing to compliance and, thus, for eventually achieving convergence.

4.2 From Peer Review Programme to Mutual Learning Programme: how did we get there?

However, the design of the peer review procedure used within the EES has been subject to some changes. The Mutual Learning Programme (MLP) that nowadays embodies the peer review idea was only created as such in 2005, after the Commission proposed substantial reforms to the EES, including an overhaul of the then-existing Peer Review Programme (PRP). The call for a reform of the employment OMC resulted from an increasing awareness that the open method of coordination had fallen short of expectations, and that delivery by the Member States needed to be ratcheted up, if the EU was to reach the 'Lisbon objective' of 'rais[ing] the employment rate from an average of 61% today to as close as possible to 70% by 2010'.¹²⁸

The Commission's review of the EES in 2002 went some way towards acknowledging the shortcomings of the employment OMC. It called for a reinforcement of the functioning of multilateral surveillance, which was regarded as being 'of considerable added value' for enhancing the Member States' delivery on their commitments, through simplifying the Employment Guidelines, developing adequate comparable statistics and indicators for assessment, and enhancing the levels of stakeholder participation.¹²⁹

The issue of improving multilateral surveillance was taken further by the 'Kok Report',¹³⁰ which emphasized that improving the process of delivery was highly

¹²⁷ With the exception of the Broad Economic Policy Guidelines (BEPGs), see Article 99 TEC.

¹²⁸ European Council, Presidency Conclusions of the Lisbon European Council, Brussels, 23-24 March, 2000, paragraph 30.

¹²⁹ Commission of the European Communities (CEC), Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. Taking Stock of Five Years of the European Employment Strategy, COM (2002) 416 final, 17 July 2002.

¹³⁰ Report of the High Level Group chaired by Wim Kok, contributing to the mid-term review of the Lisbon Strategy by identifying measures for the European economies to achieve the Lisbon goals and objectives, 2004.

contingent upon reinforcing the functioning of peer pressure and making more use of benchmarking. As central elements of the OMC, both were considered to represent ‘clear incentives for Member States to deliver on their commitments by measuring and comparing their respective performance and facilitating exchange of best practice’.¹³¹ However, some of the suggestions that followed from these considerations and called a better use of indicators, a firm implementation of the ‘naming, shaming and faming approach’ and more effective communication of the results of benchmarking to the public,¹³² were clearly too bold to be taken forward by the Commission.

Instead, the emphasis in a Commission Communication of 2005 on the need to increase the commitment on the part of both the Member States and the European institutions, mobilizing the relevant stakeholders, and simplifying the guidelines and reporting procedures was very much in line with the relatively ‘soft’ stance taken in its review of the EES in 2002. If anything, bolstering the sanctions mechanism by making full use of the ‘naming, faming and shaming approach’, as suggested in the Kok Report, was not on the shortlist.¹³³

4.3 The Mutual Learning Programme: core elements

The softer line taken by the Commission is clearly visible in the nature of the reforms proposed to improve multilateral surveillance, and in the resultant creation of the new Mutual Learning Programme (MLP). The MLP primarily aims ‘to encourage mutual learning at all levels [through the exchange of good practices and experiences] and to enhance the transferability of the most effective policies within key areas of the European Employment Strategy’.¹³⁴ In line with the Commission’s awareness that a ‘one-size-fits-all policy would be ineffective and potentially counterproductive’,¹³⁵ there is a recognition that, while Member States share common objectives and targets, one cannot presume that policy transfer is self-evident.

An assessment of whether, and how good practices can be effectively transferred to Member States is considered essential in this regard, and any such assessment must be

¹³¹ Report of the High Level Group chaired by Wim Kok, contributing to the mid-term review of the Lisbon Strategy by identifying measures for the European economies to achieve the Lisbon goals and objectives, 2004, p. 42.

¹³² Report of the High Level Group chaired by Wim Kok, contributing to the mid-term review of the Lisbon Strategy by identifying measures for the European economies to achieve the Lisbon goals and objectives, 2004, p. 42 – 43.

¹³³ Commission of the European Communities, Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Non-Discrimination and Equal Opportunities for All – A Framework Strategy, COM (2005) 224 final, 2005.

¹³⁴ Commission of the European Communities, Report on the Practical Operation of the Methodology for a Systematic and Rigorous Monitoring of Compliance with the Charter of Fundamental Rights, COM (2009) 205 final, 2009.

¹³⁵ Commission of the European Communities, Communicating to the Spring European Council. Working Together for Growth and Jobs. A New Start for the Lisbon strategy. Communication from President Barroso in Agreement with Vice-President Verheugen, COM (2005) 24, 2 February 2005, p. 28.

done by conducting evidence-based reviews of good practice in other Member States.¹³⁶ The broadly-perceived need to address both the lack of stakeholder involvement, as well as the absence of public awareness of the OMC process, was addressed by endowing the MLP with the additional objective of ‘encourage[ing] stakeholders to promote the wider and more effective dissemination of information about the [EES] and its implementation’.¹³⁷

To reach this twofold objective, the MLP framework consists of three strands of activities, under which specific relevant labour market themes are discussed.¹³⁸

1. *Thematic Review Seminars* (TRSs) are EU-level, one day meetings, which are organized twice a year and focus on a broad theme of general interest in the field of employment. Together with independent policy experts on the selected priority theme, a broad group of policymakers and stakeholders from across Europe attend the seminars. Usually, a TRS hosts approximately 150 participants.¹³⁹ The general theme discussed at the TRS will subsequently be taken forward for more detailed discussion during Peer Review meetings hosted by individual Member States (see below), and the TRS can thus be considered as having an agenda-setting role. The discussions are organized on the basis of presentations from several Member States that perform particularly well in the area that is subject to examination, or where major achievements have materialized. With the intention of stimulating the exchange of experiences and good practices in mind, the Commission takes on a leading role within the TRSs. The practical coordination and organization of the seminars is carried out on behalf of the Commission by an external contractor, which is selected through a public tendering process. The results of the TRSs are published online.
2. *Peer Reviews* (PRs) take place six times a year, are hosted by individual Member States and cover a number of single initiatives related to selected employment policy practices, in line with the priority themes of the EES. The specific objectives of PRs are; a) to identify, evaluate and disseminate good practice; b) to assess whether and how good practices can be effectively transferred to other Member States; c) to provide a learning opportunity about the implementation

¹³⁶ Commission of the European Communities, Report on the Practical Operation of the Methodology for a Systematic and Rigorous Monitoring of Compliance with the Charter of Fundamental Rights, COM (2009) 205 final, 2009.

¹³⁷ Commission of the European Communities, Report on the Practical Operation of the Methodology for a Systematic and Rigorous Monitoring of Compliance with the Charter of Fundamental Rights, COM (2009) 205 final, 2009.

¹³⁸ Commission of the European Communities, Report on the Practical Operation of the Methodology for a Systematic and Rigorous Monitoring of Compliance with the Charter of Fundamental Rights, COM (2009) 205 final, 2009.; Hardenbol, J., The Evaluation Methods of the Mutual Learning Programme of the European Employment Strategy, Paper Written for the MA Programme of European Public Affairs, Maastricht University, 2009.

¹³⁹ Participants include three or four national representatives for every Member State, representatives from the countries that are members of the European Free Trade Area (EFTA) or the European Economic Area (EEA), and delegates from the candidate countries, as well as other relevant stakeholders, such as the social partners, international organizations and further interested parties.

process or policy approaches and programmes in the field of the thematic priority; and d) to follow-up and implement the objectives of the EES. PRs are characterized by discussions in a much smaller setting than the TRSs: the participants in a PR are government representatives from a maximum of twelve interested Member States, independent experts and representatives from the European Commission. Member States that are interested in participating in a PR must apply to the Commission for a place.¹⁴⁰ Member States, in coordination with the Employment Committee (EMCO), submit examples of ‘good practices’. If EMCO confirms that the respective submission indeed constitutes a ‘good practice’, the Member State that submitted the policy example is asked to organize a PR. The organization and coordination of PRs, including the provision of practical support to the host country, is carried out by the same external contractor that is hired for the organization of the TRSs. As the agenda for the PRs is set by EMCO, the host country and the external contractor, and the meetings are essentially intended to provide an opportunity for a small group of participants (mostly experts rather than highly political officials) to discuss the details and impacts of certain policies, the Commission’s role is less prominent than under the TRSs. The results of the activities carried out under the PRs are published online.

3. Finally, *follow-up and dissemination activities* complement the European-wide mutual learning activities and engage a broader group of national stakeholders. Their aims are; a) to develop partnerships or networks which pursue the identification and exchange of good practice in a transnational context; and/or b) to contribute to mutual learning within Member States and between Member States of the most effective policies and practices, with the inclusion of all key decision makers and stakeholders; and/or c) to promote a wider and more effective dissemination of knowledge about the EES and its implementation to national and/or European-wide stakeholders. Member States may apply for funding for such projects to the Commission, which will then evaluate the proposals and decide on whether to grant a subsidy to the respective applicant(s) of up to 80% of the budgeted project costs.

4.4 Assessing the effectiveness of the MLP

Compared to the earlier Peer Review Programme, the potential of the Mutual Learning Programme to bring about effective policy learning seems considerably greater – it demonstrates a strengthened focus on mutual learning, by enhancing the opportunities to exchange experiences and good practices amongst cross-national policymakers and experts, during what seems to be a context sensitive, lesson drawing approach. In addition, it engages a broader range of stakeholders from different territorial levels and from different societal sections across Europe.

¹⁴⁰ Experience to date indicates that the Commission usually does not reject applications, and that it allows the number of interested Member States to exceed the number of seats available. However, countries that are late in requesting participation in a PR are unable to participate.

However, how much of this potential has been realized? We now examine the extent to which the MLP may be considered positive in bringing about learning, and stimulating compliance with the EES guidelines and recommendations, and investigate the factors inherent or related to the MLP that have clearly hindered such progress.¹⁴¹

1. Cognitive shifts... but policy transfer?

The general perception of the effectiveness of the EES for bringing about learning and policy transfer seems quite apparent: the impact of the Mutual Learning Programme must be considered as being largely cognitive. Indeed, the success of the MLP derives largely from the notion that it has been an effective mechanism for identifying common European challenges and promising policy approaches – which have in turn contributed to broad shifts in national policy thinking – and for enhancing the mutual awareness of policies, practices and problems in other Member States. In addition, the programme has been credited with pushing the Member States to rethink their established approaches and practices, as a result of comparisons with other countries on the one hand, and the obligation to re-examine and re-evaluate their own policies and performance on the other.¹⁴²

The largely cognitive nature of the MLP implies that its effectiveness for bringing about concrete policy transfer at the national level must be considered as being more limited. As a matter of fact, there seem to have been only a few instances of direct policy transfer,

¹⁴¹ Before moving on to this examination, it is important to acknowledge the limitations inherent in any evaluation that purports to assess the effectiveness of mutual learning processes. If any attempt to evaluate such notions as the ‘transfer of ideas’ or the ‘sharing of knowledge’ already poses inherent difficulties related to measurement, we also note that it is difficult to establish the existence of a causal relationship between the MLP and any observed cognitive shifts and/or concrete policy changes. This is linked to recognition that changes in Member States’ priorities and/or policies may not gain their inspiration from mutual learning processes, but actually precede the launch of such processes. Furthermore, we must keep in mind that Member States may both try to ‘upload’ their domestic priorities and preferences to the supranational level as well as be selective in what they ‘download’ from the European level. Related to this is the notion that, for strategic reasons, governments may consciously downplay or overstate the influence of the EES on domestic policymaking processes, both in their communication to domestic audiences, as well as in their reporting procedures to the EU. In short, any evaluation of the MLP should be guided by a realization that the relationship between OMC processes and Member State policies should be understood as ‘a two-way interaction rather than a one-way causal impact’. See Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 453-57.

¹⁴² See for example Casey, B. and Gold, M., *Peer Review of Labour Market Programmes in the European Union: What can Countries Really Learn from One Another?*, *Journal of European Public Policy*, 2005, p. 12, 23-24.; Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 46, 497-532.; López-Santana, M. *The Domestic Implications of European Soft Law: Framing and Transmitting Change in Employment Policy*, *Journal of European Public Policy*, 2006, p. 13, 481 - 499.; Van Vliet, O. and Koster, F., *Europeanisation and the Political Economy of Active Labour Market Policies*, Paper prepared for the NIG Annual Work Conference 2008, Enschede, 20-21 November, 2008.; Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005.

an observation which has led the majority of scholars to conclude that, whilst the peer review process may have established a learning process, its practical impact has been limited.¹⁴³ In other words, there seems to be a widely held view that peer review:

hardly acted as a catalyst for policy transfer. Indeed, it might even have taught participants that transfer is seldom an easy option. It established an opportunity for ideas, suggestions and observations to be exchanged, and for the validity and appropriateness of particular policy measures in particular policy contexts, to be assessed [but] it contributed not so much to ‘emulation’, ‘harmonization’, or ‘penetration’ of policies.¹⁴⁴

Although these views are based on an evaluation of the earlier Peer Review Programme, their validity does not seem to be compromised under the new Mutual Learning Programme. As a matter of fact, a synthesis report evaluating the activities held during the second semester of the MLP in spring 2006, confirms their continued validity, by suggesting that:¹⁴⁵

[t]he mutual learning programme ... offered a rich menu of interesting and stimulating examples of practices and reforms... There was considerable potential for learning and while there was agreement on many fundamentals, participants often raised questions concerning the transferability of practices or reforms.¹⁴⁶

The quotations clearly confirm the findings of the academic literature related to the effectiveness of the Peer Review Programme for bringing about learning and policy transfer: while the potential for the MLP to bring about cognitive changes should not be underestimated, at the same time it is clear that direct policy transfer is not to be regarded as self-evident. The awareness that both the variety of domestic institutional structures, as well as the different specific challenges faced by the Member States, render any form of direct policy transfer virtually impractical, is clearly in line with the idea put forward in the academic literature that it would be ineffective, or even counterproductive, to assume that, in a situation of great transnational administrative, institutional and political differences, a ‘one-size-fits-all approach’ will suffice. The fact that policymakers and experts are conscious of this must be considered positive, as it allows for the

¹⁴³ See, for example, Casey, B. and Gold, M., Peer Review of Labour Market Programmes in the European Union: What can Countries Really Learn from One Another?, *Journal of European Public Policy*, 2005, p. 36.; Zeitlin, J., The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 472.

¹⁴⁴ Casey, B. and Gold, M., Peer Review of Labour Market Programmes in the European Union: What can Countries Really Learn from One Another?, *Journal of European Public Policy*, 2005, p. 36 – 37.

¹⁴⁵ This observation is based on an examination of several synthesis reports and the Commission’s annual report assessing the functioning and usefulness of the MLP.

¹⁴⁶ Visser, J. and Kouw, M., Synthesis Report. Ensuring Effective Implementation through Better Governance, 2006. Available at: http://pdf.mutual-learning-employment.net/pdf/synthesisReports/SynthesisReportSpring2006_EN.pdf.

development of a context sensitive, lesson drawing approach, as a result of which policy learning may take place more effectively and efficiently.¹⁴⁷

2. Stakeholder buy-in: improvements on paper, setback in practice?

The Peer Review Programme that existed before the 2005 overhaul of the Lisbon Strategy was heavily criticised for its exclusive nature. The perception that the formulation of the Employment Guidelines, common indicators for assessment, the Joint Employment Report and the recommendations to the Member States, was undertaken primarily by experts in the Employment Committee, resulted in the suggestion that, rather than coming close to the ideal-type of participatory governance promulgated at Lisbon, the peer review process may be more properly considered as a learning process for a limited community of labour market technicians and experts'.¹⁴⁸ Several sources confirm the idea that, particularly in the field of employment policies, 'the interest of NGOs beyond the social partners to be involved was rather limited', and that instead of an open participation, the OMC created 'procedures and rules for an 'elite deliberation' by civil servants and experts'.¹⁴⁹

Even if this type of 'expertocratic deliberation' may contribute to the creation of a more knowledgeable community, the exclusive nature of such deliberation does not particularly contribute to enhancing bottom-up learning and, moreover, it does not create the domestic awareness of the process that is so important for making Member States deliver on their commitments. The 2005 overhaul of the Peer Review Programme therefore sought to address this issue of exclusiveness by creating new opportunities for the active engagement of a broader range of stakeholders, including representatives from local, regional and national authorities across Europe, as well as the social partners, international organisations and other interested parties.

If a quick glance at the design of the Mutual Learning Programme suggests that the opportunities for stakeholders to participate in the process have indeed been considerably broadened – the TRS and, in particular, the follow-up and dissemination activities provide ample opportunities for active stakeholder engagement in the process – there are reasons to be sceptical about the actual involvement of, and potential influence exerted by, such stakeholders in practice. One of the reasons for this is that the political nature of the TRSs limits both the probability of a genuine and effective exchange of ideas, as well as the scope for stakeholders to play an active and effective role in raising concrete policy challenges and possible ways of addressing them. This is confirmed by the 2008 annual report that evaluates the MLP, according to which approximately 80% of the participants

¹⁴⁷ Radaelli, C., *The Open Method of Coordination: A New Governance Architecture for the European Union?*, SIEPS Reports, Nr.1, 2003.; Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels: Peter Lang, 2005.

¹⁴⁸ Casey, B. and Gold, M., *Peer Review of Labour Market Programmes in the European Union: What can Countries Really Learn from One Another?*, *Journal of European Public Policy*, 2005, p. 37.

¹⁴⁹ Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 503.

found the TRSs useful, but fierce criticism was voiced regarding the general nature of the presentations and the timid debate afterwards,¹⁵⁰ both of which may relate to the domination of such meetings by government representatives acting according to national mandates.

Another reason for scepticism is that, although the opportunity for Member States to organize specific follow-up and dissemination activities involving a great variety of stakeholders must be considered theoretically promising for enhancing bottom-up learning and stimulating compliance, only a limited number of projects have thus far been implemented in practice. This clearly suggests that the potential of this tool is far from fully realised. The reasons for this are possibly linked to the fact that the Commission's call for proposals is not loud enough and/or that the Community budget 'merely' finances up to 80% of approved projects, which means that, in particular, new Member States may turn down any invitations to apply.¹⁵¹

Thus, the mutual learning process (still) falls short of the ideal-type of participatory governance promulgated at Lisbon. As a consequence of the lack of effective stakeholder input, as well as the resultant virtual absence of public awareness at the national and sub-national level, the potential of the MLP for stimulating policy learning, and enhancing compliance by the Member States with guidelines and recommendations, is not fully realised.

3. Indicators, reporting and recommendations: how politically strategic behaviour limits learning and compliance

The absence of participatory governance, as well as the overall lack of a genuine political commitment to the Luxembourg Process, particularly on the part of the Member States, are both factors that significantly hinder the effectiveness of the Mutual Learning Programme.¹⁵² Evaluations regarding the functioning of the earlier Peer Review Programme demonstrate that the lack, or at least uneven degree, of commitment on the part of the Member States, has found its expression in a number of procedural limitations that are restrict the opportunity for peer pressure and benchmarking to function adequately.

¹⁵⁰ Commission expert in Hardenbol, J., The Evaluation Methods of the Mutual Learning Programme of the European Employment Strategy, Paper written for the MA Programme of European Public Affairs, Maastricht University, 2009.

¹⁵¹ Hardenbol, J., The Evaluation Methods of the Mutual Learning Programme of the European Employment Strategy, Paper written for the MA Programme of European Public Affairs, Maastricht University, 2009, p. 12 – 13.

¹⁵² Commission of the European Communities, Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Non-Discrimination and Equal Opportunities for All – A Framework Strategy, COM (2005) 224 final, 2005, p. 14.; Report of the High Level Group chaired by Wim Kok, contributing to the mid-term review of the Lisbon Strategy by identifying measures for the European economies to achieve the Lisbon goals and objectives. 2004, p. 39, 44 – 45; Zeitlin, J., The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy, ed Pochet, P. and Zeitlin, J., The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies, Brussels, Peter Lang, 2005, p. 492 – 493.

One of the observed restrictions relates to the limited comparability¹⁵³ of national structures and processes, which is mostly due to the sparing use, or even avoidance, of quantitative indicators and the questionable relevance of any causal significance such indicators may exhibit.¹⁵⁴ The persistent absence of comparable data partly originates from the fact that the interest of Member States to genuinely learn from each other is circumscribed by their primary interest to avoid negative evaluations of their national situation, in order not to offer any domestic opposition, the media or the public, grounds for voicing criticism.

The focus of national representatives on limiting the comparability of national policies and processes by avoiding the use of precise indicators is clearly apparent from the following statement from a member of the Employment Committee, who argues that:

[t]he discussions focus often on the avoidance of quantifiable indicators, as through this the Member States are more easily comparable. Ideally, there would be clear objectives. The Commission has proposed considerably more precise, quantified indicators. The Member States are, however, very hesitant when it comes to precise objectives because the intensity of political pressure also depends on these indicators. Then the minister asks: “Why are we ranked last?”¹⁵⁵

If improving the use of indicators was recognised by the European Commission as an important element for stepping up the peer review process,¹⁵⁶ it seems that the actual reforms to the Luxembourg Process have been modest and are certainly not sufficient to

¹⁵³ ‘Comparability’ is here understood to mean that ‘data (estimates) for different populations (whether countries or different groups within the same country) can be legitimately (i.e. in a statistically valid way) put together (aggregated), compared (differenced), and interpreted (given meaning) in relation to each other and against some common standards’: see Verma, V., *Issues in data quality and comparability in EU-SILC*, 2006, p. 10. Comparability of data is particularly important, as this provides the basis on which EU-wide effective benchmarking and the definition of best practices can take place (idem.).

¹⁵⁴ Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 511.; Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 475.

¹⁵⁵ Member of the Employment Committee in Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 511. Although the lack of adequately defined indicators clearly constitutes a significant impediment for achieving full comparability, there are also other sources of non-comparability - reporting errors, differences in data treatment, different definitions of measured units and differences regarding data collection methods, to mention a few - which all constitute potential impediments for arriving at full inter-country comparability. Drawing on Verma (Verma, V., *Issues in data quality and comparability in EU-SILC*, 2006, p. 10), we conclude that there is a clear correlation between the extent of diversity of methodology and implementation arrangements on the one hand, and the complexity and acuteness of the problem of (non-)comparability on the other.

¹⁵⁶ Commission of the European Communities, *Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Non-Discrimination and Equal Opportunities for All – A Framework Strategy*, COM (2005) 224 final, 2005, p. 16.

maximize the opportunities for mutual learning.¹⁵⁷ An examination of the currently applicable Employment Guidelines clearly highlights the ongoing lack of comparable and precisely defined indicators that are ‘outcome-oriented, responsive to policy interventions, subject to a clear and accepted normative interpretation, timely, and revisable’, and which could hence serve as tools for learning and self-corrective action by national and local actors.¹⁵⁸

While this continued and less than optimal use of indicators under the Mutual Learning Programme limits the effectiveness of the programme for bringing about learning and compliance, the reforms brought about to simplify national reporting procedures¹⁵⁹ appear to have failed to achieve the goal of facilitating peer review and ensuring the comparability of national programmes. As a matter of fact, it is argued that replacement of the National Action Plans for employment (NAPs/empl) with National Reform Programmes (NRPs), in line with the 2005 architectural shift to issue Integrated Guidelines for Growth and Jobs rather than separate guidelines in the economic and employment fields, has led to ‘greater unevenness in Member States employment policy reporting and a loss of European-level monitoring capacity’.¹⁶⁰

It seems that the majority of Member States do not structure their NRPs around the Integrated Guidelines: only some Member States include full data on the existing European employment indicators, and allegedly, few report progress on the EES targets.¹⁶¹ It is not difficult to imagine that this selectivity in reporting practices by Member States may be greatly motivated by political resistance to the comparability of national programmes, as described above.

Finally, the strategic behaviour of Member States is visible not only in their cautious approach to the establishment of common indicators and uneven reporting practices to the EU, but also from the attempts of national representatives to avoid or modify recommendations formulated by the Commission. Given that ‘a crucial objective of national officials in the bilateral negotiations with the Commission [where Member States are given the opportunity to elucidate aspects of their NRPs] is the adoption of the

¹⁵⁷ Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, La Follette School Working Paper Series, No. 2007-031, 2007, p. 14.

¹⁵⁸ Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, La Follette School Working Paper Series, No. 2007-031, 2007, p. 14.

¹⁵⁹ Commission of the European Communities, Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Non-Discrimination and Equal Opportunities for All – A Framework Strategy, COM (2005) 224 final. 2005, p. 31-32.

¹⁶⁰ Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, La Follette School Working Paper Series, No. 2007-031, 2007, p. 8.

¹⁶¹ Zeitlin, J., A Decade of Innovation in EU Governance: The European Employment Strategy, the Open Method of Coordination and the Lisbon Strategy, La Follette School Working Paper Series, No. 2007-031, 2007, p. 8.

most ‘government-compatible’ formulation of these recommendations’,¹⁶² the outcome of the evaluation process – namely, the recommendations that are eventually adopted for each Member State – result from what is, in fact, a political bargaining process aimed not so much at arriving at effective and objective outcomes, but at avoiding any critical evaluation of the domestic situation. Although Member States may be more easily pushed to comply with such weakened recommendations, their utility in bringing about real and necessary policy change is questionable.

4. Between sanctions and learning: how to make Member States comply?

Perhaps due to the inherently political and ‘soft’ nature of such recommendations, some authors find evidence that the cognitive influence of the Mutual Learning Programme should be considered more effective for bringing about policy change than the normative influence of Council recommendations.¹⁶³ Similar ideas are shared by a number of other authors, who find that, even if there are no tangible sanctions from the supranational level, we have seen an ‘indirect coercive transfer’ of policies from the European to the national level, as a result of the ‘framing effect’ produced by the mutual learning process. Given the (assumed) significant potential of the cognitive dimension to bring about change, these authors suggest that sanctions may not be the key for bringing about successful compliance.¹⁶⁴

However, others argue that there is evidence that ‘peer pressure and associated practices such as recommendations and rankings do have an influence on the behaviour of Member State governments’, provided that the recommendations are perceived as legitimate by the Member State, the Member State is sensitive to criticism from the EU, and the EES process enjoys a certain degree of visibility at the domestic level.¹⁶⁵ These authors argue that, in many cases, Member States have been prepared to take some corrective action to avoid EU recommendations and low rankings on common indicators.¹⁶⁶ In line with this point of view, the continued absence of an effective ‘naming, faming and shaming

¹⁶² Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 511., cf. Casey, B. and Gold, M., *Peer Review of Labour Market Programmes in the European Union: What can Countries Really Learn from One Another?*, *Journal of European Public Policy*, 2005, p. 25.

¹⁶³ Van Vliet, O. and Koster, F., *Europeanisation and the Political Economy of Active Labour Market Policies*, Paper prepared for the NIG Annual Work Conference 2008, Enschede, 20-21 November, 2008, p. 18.

¹⁶⁴ López-Santana, M., *The Domestic Implications of European Soft Law: Framing and Transmitting Change in Employment Policy*, *Journal of European Public Policy*, 2006, p. 495; Report of the High Level Group chaired by Wim Kok, contributing to the mid-term review of the Lisbon Strategy by identifying measures for the European economies to achieve the Lisbon goals and objectives, 2004.

¹⁶⁴ Radaelli, C., *The Open Method of Coordination: A New Governance Architecture for the European Union?*, SIEPS Reports, Nr.1, 2003; Trubek, D. and Trubek, L., *Hard and Soft Law in the Construction of Social Europe: The Role of the Open Method of Coordination*, *European Law Journal*, 2005.

¹⁶⁵ Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 477.

¹⁶⁶ Zeitlin, J., *The Open Method of Communication in Action. Theoretical Promise, Empirical Realities, Reform Strategy*, ed Pochet, P. and Zeitlin, J., *The Open Method of Coordination in Action. The European Employment and Social Inclusion Strategies*, Brussels, Peter Lang, 2005, p. 477.

approach’ implies that the potential for the Luxembourg Process to stimulate compliance of Member States with the Employment Guidelines and country-specific recommendations is as yet not fully exploited.

Other scholars perceive the current lack of ‘hard’ sanctions for those who fail to follow up recommendations or meet targets as an explanation for the limited degree of policy convergence, as well as being problematic for bringing about compliance.¹⁶⁷ Neither the cognitive influence exerted by the Mutual Learning Programme, nor possible soft sanctions, such as publicity in the media or even ‘naming-and-shaming’, are considered powerful enough to make Member States comply with guidelines and/or recommendations. Following this line of thinking, one might argue that the weakly developed strategic dimension of the MLP, involving few positive incentives (for example, financing for projects) and an absolute lack of negative sanctions (for example, fines), constitutes an important factor hindering compliance with guidelines and/or recommendations.

4.5 Lessons learned: how to improve the effectiveness of the MLP?

Although the above discussion may not be conclusive as to whether certain factors stimulate or hinder compliance, it does allow for some valuable lessons. The suggestions for improving the functioning of the OMC evaluation mechanism provided below are inspired not only by our evaluation of the peer review process as used in the EU, but also by the lessons learned from the functioning of the peer review mechanism in the context of the OECD, as touched upon in the previous chapter. What lessons can be learned from both discussions? What concrete suggestions can we put forward for improving the effectiveness of the MLP or, in fact, any evaluation mechanism in which mutual learning takes on a centre-stage position? Specifically, it seems that the effectiveness of any mutual learning process is highly contingent upon the following factors:

1. **The iteration of the process.** The repeated nature of the process is nothing less than a prerequisite both for stimulating cognitive shifts amongst policymakers, as well as for strengthening the impact of peer pressure, which may act as an incentive for Member States to deliver on their commitments;
2. **The recognition of cross-national differences** regarding domestic institutions, administrative structures and political cultures. A context sensitive approach characterised by a realization that national policy approaches do not have to be identical for convergence towards certain goals to be arrived at, and that a ‘one-size-fits-all approach’ may not suffice, enhances the effectiveness and efficiency of mutual learning and may increase the ‘adapted transfer’ of ‘good practices’ in a policy mix that best suits each Member State;

¹⁶⁷ Heidenreich, M. and Bischoff, G., *The Open Method of Coordination: A Way to the Europeanization of Social and Employment Policies?*, *Journal of Common Market Studies*, 2008, p. 524 – 525; Kerber, W. and Eckardt, M., *Policy Learning in Europe: The Open Method of Coordination and Laboratory Federalism*, *Journal of European Public Policy*, 2008, p. 236.

3. **The active engagement of a broad range of stakeholders**, including the social partners, civil society organizations, local and regional authorities, and national parliaments. The meaningful involvement of relevant actors is deemed essential for enhancing (bottom-up) learning and, through the visibility that goes along with it, for increasing the delivery by Member States on their commitments;
4. **The existence of common European, quantitative indicators of assessment** that are sufficiently comparable and disaggregated to serve as diagnostic tools for self-corrective action by both local and national actors. The establishment of precisely defined and explicit indicators that are at the same time both revisable and diagnostic is crucial for effective monitoring and peer review;
5. **Uniform and consistent reporting by Member States** on their progress towards each objective and guideline. This is important for achieving comparability of national policies, which in turn is a crucial factor for enabling effective EU level monitoring, for enhancing peer pressure and, thus, for increasing mutual learning and compliance on the part of the Member States;
6. **The unconditional acceptance by Member States of objective – and, if need be, firm – recommendations**. Only if Member States accept recommendations that objectively identify problem areas and the necessary steps to be taken will compliance lead to meaningful policy improvements;
7. **The introduction of appropriate, context related incentives stimulating compliance**. Given the lack of academic consensus on the virtues of cognitive learning versus the merits of ‘hard’ sanctions, it seems difficult to provide any concrete suggestions on this issue. The main reason for this lack of consensus lies in the variation in ‘regulatory’ or ‘legal’ cultures in all 27 Member States (see also the Annex to this report - constructivist mechanisms). We must therefore keep in mind that Member States react differently to cognitive influences, peer pressure, ‘naming-and-shaming exercises’, and even to sanctions;
8. Finally, a **genuine commitment on the part of the Member States** to make the evaluation mechanism work. A sustained political determination of the highest order is the principal precondition for effecting any real improvements to the evaluation mechanism, and for arriving at truthful compliance with its outcomes.

4.6 Lessons confirmed – Part I: evaluation mechanisms in the anti-discrimination field

If the recommendations proposed above are primarily based on a very specific analysis addressing the effectiveness of the MLP for stimulating compliance with the European Employment Guidelines, their wider validity appears to be confirmed by an examination of other policy areas where evaluation mechanisms are employed. Indeed, even a quick glance at the design, and functioning in practice, of such evaluation mechanisms, suggests that, very often, similar issues arise that either hinder the functioning of the

evaluation mechanism, or should be regarded as (potentially) stimulating compliance by Member States, within both open and more rule-based integration frameworks.

In order to substantiate this statement, the focus of this section of the report will be on the functioning of the evaluation mechanisms employed for ensuring Member States' compliance with the Racial and Employment Equality Directives.¹⁶⁸ The analysis below clearly suggests that, even if the integration framework in this case differs markedly from the one discussed in the main part of this chapter – while the latter is characterised by 'soft' guidelines and voluntary actions by Member States, the former clearly encompasses binding rules that require correct implementation by the Member States – the twin notions of peer review and mutual learning involving a wide range of stakeholders take on a centre-stage position, and must be considered useful for enhancing compliance within this field as well.

Another – less fortunate – similarity is that issues similar to those identified as hindering the full realisation of the Mutual Learning Programme also seriously diminish the potential of the evaluation mechanisms to stimulate compliance within the anti-discrimination field. However, apart from these notable similarities, the analysis below paints a picture that is sometimes markedly – and often positively – different from what we have identified to be the case for the Mutual Learning Programme. However, for reasons that will become apparent from this analysis, rather than disqualifying our recommendations in the previous section, these differences further support their validity.

1. Reporting, sanctioning and data collection issues

We will first consider one factor that significantly – and positively – diverges from what we have seen to be the case for the Mutual Learning Programme: the procedures for reporting on the performance of the Member States. If the prevalence of self-reporting by the Member States within the framework of the MLP was identified as a less than optimal way of informing the European Commission about national performances, the notion of reporting by independent experts to the Commission in the anti-discrimination field greatly benefits the effective evaluation of compliance with the anti-discrimination directives and, hence, also the Commission's ability to effectively and efficiently fulfil its function as 'guardian of the Treaties'.

Independent reporting to the European Commission may, in fact, originate from three broad sources. First of all, the European Network of Legal Experts in the anti-discrimination field provides the Commission with 'independent information and advice on implementation and application of the two anti-discrimination directives'.¹⁶⁹ Obviously, reporting by **high-level, independent experts** is more likely to produce **objective and impartial information** on how the provisions of the directives have been

¹⁶⁸ Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, and Council Directive 2000/73/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation.

¹⁶⁹ European Network of Legal Experts in the Non-Discrimination Field, *European Anti-Discrimination Law Review*, No. 2, 2005, p. 7.

incorporated into national law than when Member States themselves report on their own performance. The probability that facts are consciously and purposefully misrepresented – something which, for obvious political reasons, appears not to be exceptional in cases where self-reporting by Member States is the rule¹⁷⁰ – is significantly reduced.

Even if Member States are given the opportunity to read and comment on the country reports produced by the Network,¹⁷¹ the ‘expertocratic’ procedure for reporting on the implementation of the anti-discrimination directives comes quite close to fulfilling the independency requirement that is essential for enabling an evaluation mechanism to function effectively: it provides a sound basis for producing the unbiased information, according to which the Commission may verify compliance efficiently and, if need be, decide to launch **infringement proceedings**.

At this point, it should be noted that the **standard template** according to which the national experts report to the Commission, is highly valuable, as it makes it easy for the Commission to understand the relative performances of the Member States and to prepare the **comparative analyses** on the Member States’ progress that are made **publicly available** by the Network.¹⁷² The transparency that this provides in relation to the Member States’ relative performance implies that Member States are *de facto* **ranked** and that **‘naming-and-shaming’** may act as a soft sanctioning mechanism, next to the Commission’s use of hard sanctions where it (successfully) launches infringement proceedings.¹⁷³

A second source of independent information is found in the EU Fundamental Rights Agency. Thanks to its task of collecting and analysing ‘objective, reliable and comparable information on the development of the situation of fundamental rights’, and its right to formulate **opinions** to the Union’s institutions, or to the Member States, on the implementation of the anti-discrimination directives,¹⁷⁴ the Fundamental Rights Agency could clearly act as another useful provider of independent information and expertise to the European Commission. Its current study on the impact on the ground of the Racial Equality Directive, which will be a contribution to the Commission’s report on the application of this Directive,¹⁷⁵ is a good illustration of how the Commission may harness the expertise of the Agency to support its own function as ‘guardian of the Treaties’. However, the absence of any comprehensive evaluation report on the functioning of this

¹⁷⁰ This is, for example, clearly the case for the MLP – please refer to the first part of this chapter.

¹⁷¹ European Network of Legal Experts in the Non-Discrimination Field, European Anti-Discrimination Law Review, No. 3, 2006, p. 9.

¹⁷² European Network of Legal Experts in the Non-Discrimination Field, European Anti-Discrimination Law Review, No. 4, 2006, p. 9.

¹⁷³ In the preceding sections of this chapter, it was argued that ranking and ‘naming-and-shaming’ may be effective tools for stimulating compliance, depending on whether the Member States consider the EU-level process to be legitimate and are sensitive to criticism from the EU, and also on whether the evaluated legislation enjoys some degree of visibility at the national level.

¹⁷⁴ European Union Fundamental Rights Agency, 2007, p. 46.

¹⁷⁵ Commission of the European Communities, Mutual Learning Programme of the European Employment Strategy, 2009, p. 8. Available at <http://www.mutual-learning-employment.net/>.

newly established Agency means that it is premature to draw any firm conclusions regarding its effectiveness in practice.

Thirdly, since they are endowed with the tasks of ‘conducting independent surveys’ and ‘publishing independent reports and making **recommendations**’ on any issue related to discrimination,¹⁷⁶ National Equality Bodies also have an important role in stimulating compliance with the anti-discrimination directives. However, even though the latest comprehensive evaluation report on the functioning of these Bodies concluded that it was too early to make a firm assessment on their effectiveness, it does suggest that there may be some reasons for concern, in that the theoretical potential of these Bodies to support the Commission has not been fully exploited in practice.

It is, for example, pointed out that ‘confusion about what is meant by . . . conducting surveys and issuing reports and recommendations’ and ‘insufficient funding and a lack of (well trained) staff’ have both resulted in ‘very few surveys’ and ‘very few reports and recommendations’ being published. Other concerns relate to the degree to which Equality Bodies are able to carry out their mandate free from undue governmental interference.¹⁷⁷

Thus, any preliminary conclusions regarding the effectiveness of this potentially promising mechanism should perhaps focus on the importance of creating the appropriate circumstances for **truly independent reporting** to take place, and on endowing the Equality Bodies with the **capacity and resources** to actually function in the way they are expected.

A final point in this section relates to the issue of data collection, which leads to problems similar to those identified for the Mutual Learning Programme. There seems to be a widely held perception that there is a wide variability of data from the Member States, which makes it very difficult, ‘if not impossible’, to achieve the degree of data comparability that is essential for assessing the scale and nature of discrimination suffered, and for effectively monitoring and evaluating policies.¹⁷⁸

¹⁷⁶ Directive 43/2000 EC, Article 13(2).

¹⁷⁷ Bell, M., *The Implementation of European Anti-Discrimination Directives: Converging Towards a Common Model?*, *The Political Quarterly*, 2008, p. 40 - 42; Holtmaat, R., *Catalysts for Change? Equality Bodies According to Directive 2000/43 EC – Existence, Independence and Effectiveness*, Luxembourg, Office for Official Publications of the European Communities, 2006, p. 5 – 7; European Network of Legal Experts in the Non-Discrimination Field, *European Anti-Discrimination Law Review*, No. 1, 2005, p. 23 - 28.

¹⁷⁸ Commission of the European Communities, *Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Non-Discrimination and Equal Opportunities for All – A Framework Strategy*, COM (2005) 224 final, 2005, p. 4; Commission of the European Communities, *Report from the Commission to the Council and the European Parliament. The Application of Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin*, COM (2006) 632 final 2, 2006, p. 9.; Commission of the European Communities, *Communication From the Commission to the Council and the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Non-Discrimination and Equal Opportunities: A Renewed Commitment*, COM (2008) 420 final, 2008, p. 7; Makkonen, T., *Measuring Discrimination, Data Collection and EU Equality Law*, Luxembourg: Office for Official Publications of the European Communities, 2007.

Whereas the lack of truly comparable data within the MLP process may originate from the individual Member State's strategic behaviour to be selective in the presentation of data, the uneven existence of data in the anti-discrimination field seems mostly related to the fact that the majority of countries do not have a tradition of collecting data related to discrimination.¹⁷⁹ If the relatively small body of data available in this area is currently to be regarded as problematic, it may also be considered advantageous, at least in the sense that it gives rise to ample opportunities for **streamlining and standardizing methods of data collection**, something that is essential if the situation within, and the performance of, Member States is to be effectively monitored.¹⁸⁰

2. Stakeholder involvement: promising on paper, disappointing in practice

While a consideration of comparable data and impartial reports are an essential precondition for the Commission to be able to effectively function as 'guardian of the Treaties', there is – just as we have seen in our analysis regarding the Mutual Learning Programme – also recognition of the added value of the **active engagement of the social partners and non-governmental organisations** for the promotion of proper implementation of EU legislation. The anti-discrimination directives oblige the Member States to inform the relevant stakeholders of the provisions of the directives and to promote a dialogue both between the social partners, and with the relevant non-governmental organisations.¹⁸¹

These legal provisions indicate a widely shared recognition that active stakeholder involvement is essential for making Member States comply. However, an examination of the extent to which they have been effectively implemented in practice appears to present 'a mixed picture'.¹⁸² As a result, the degree to which stakeholders are actively engaged is disappointingly low.

The Commission's reaction to the low degree of awareness amongst stakeholders, and the correspondingly low extent to which they are involved, has been to undertake several **complementary measures** to ensure that anti-discrimination legislation and the Member States' (non-)compliance with this legislation are firmly brought to the attention of the relevant actors. One such complementary measure concerns the organization by the Commission of **awareness raising activities** involving a broad range of both

¹⁷⁹ European Network of Legal Experts in the Non-Discrimination Field, European Anti-Discrimination Law Review, No. 4, 2006, p. 6.

¹⁸⁰ As a matter of fact, the Commission is now working together with the Member States' statistical authorities and the National Equality Bodies, with an eye to collecting statistics and developing a system for gathering information, respectively. See Commission of the European Communities (CEC), 25th Annual Report from the Commission on Monitoring the Application of Community Law, COM (2008) 777, Brussels, 2008, p. 7.

¹⁸¹ See Articles 10 to 12 of Directive 43/2000/EC, and Articles 12 to 14 of Directive 78/2000/EC.

¹⁸² Commission of the European Communities, Report from the Commission to the Council and the European Parliament, The Application of Directive 2000/43/EC of 29 June, 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, COM (2006) 643 final 2, 2006, p. 6 -7.

governmental as well as non-governmental actors. The establishment of 2007 as European Year of Equal Opportunities for All, the organisation of annual, high-level Equality Summits involving a broad spectrum of relevant stakeholders, and the EU information campaign ‘For Diversity – Against Discrimination’ are clear examples of this type of activities.¹⁸³

Being endowed with the task of informing the general public about their fundamental rights and ways of enforcing them, the Fundamental Rights Agency also fulfils a potentially important awareness raising role.¹⁸⁴ Finally, **training activities** organised by the Commission and targeting a range of key stakeholders also constitute a tool for disseminating information, which may contribute to achieving the necessary awareness for meaningful stakeholder involvement to be achieved.¹⁸⁵

3. Mutual learning: achieving uniformity while respecting diversity?

The previous section of this report suggests that non-governmental stakeholders are appreciated for the ‘watchdog function’ they may fulfil. However, their active involvement is deemed equally important where **mutual learning** exercises take place. In fact, mutual learning – through **peer review** and the **exchange of best practices** – is an important tool for stimulating policy change, even in a rule based framework such as the anti-discrimination field.¹⁸⁶ Within this field, learning may take place through a diversity of reports published by the Network of Legal Experts, by the Fundamental Rights Agency and by the National Equality Bodies, although the latter have thus far not been particularly active in publishing reports.

¹⁸³ Commission of the European Communities, Communication From the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Non-Discrimination and Equal Opportunities for All – A Framework Strategy, COM (2005) 224 final, 2005, p. 9; Commission of the European Communities, Communication From the Commission to the Council and the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Non-Discrimination and Equal Opportunities: A Renewed Commitment, COM (2008) 420 final, 2008, p. 8 – 9.

¹⁸⁴ European Network of Legal Experts in the Non-Discrimination Field, European Anti-Discrimination Law Review, No. 5, 2007, p. 46.

¹⁸⁵ Commission of the European Communities, Communication From the Commission to the Council and the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Non-Discrimination and Equal Opportunities: A Renewed Commitment, COM (2008) 420 final, 2008, p. 8.

¹⁸⁶ It is interesting to see that peer review exercises have been introduced not only in ‘loose integration frameworks’ such as the EES, but also in a number of policy fields that are characterized by binding rules. In fact, it seems that the relative success of the peer review mechanism in the Mutual Learning Programme has triggered the enthusiasm of several of the Commission’s Directorates-General to introduce similar mechanisms in ‘their’ areas of competence. For example, in the reformed Stability and Growth Pact, ‘peer review, peer support and peer pressure are now fully playing a key supporting role in ensuring that member States [*sic*] comply with the principles and rules of the Pact’. See Bertoldi, M., Peer Review, Peer Support and Peer Pressure: The European Union Experience, in OECD, 2008, p. 113 – 114. Thus, it appears that even in those fields where there seems little room for peer review activities, peer support and peer pressure are considered to be indispensable for ensuring proper and effective implementation of the rules.

By contrast, the frequently published country reports by the Network of Legal Experts constitute a rich source for learning: referring both to successful examples of transposition, as well as to the principal shortcomings, they could be useful sources for civil servants to learn about problems experienced by other Member States and, more importantly, about ways to successfully address them. Other publications, such as thematic reports published both by the Network and by the Agency, also offer ample opportunities for involved stakeholders to learn, by discussing specific themes in detail and highlighting best practices.¹⁸⁷

However, mutual learning opportunities are not restricted to mere ‘paper exercises’. Events such as the Year of Equal Opportunities for All, annual Equality Summits and legal seminars constitute **networking and cooperation forums, where experiences are exchanged and good practices shared** amongst EU and national policymakers, legal experts, representatives from the National Equality Bodies and non-governmental organisations. ‘Equinet’, the European network of equality bodies, was created as another forum where transnational policy learning and an exchange of experiences could take place.¹⁸⁸ Finally, the non-discrimination governmental expert group, established by the Commission to examine the impact of non-discrimination measures, to validate good practice through **peer learning** and to develop benchmarks to evaluate the effectiveness of non-discrimination policies, constitutes another useful platform in this regard.¹⁸⁹

Given that the results of learning become visible only in the medium to long term, and that the scholarly discussions to date on the extent to which actual policy change may follow from learning are inconclusive,¹⁹⁰ it is too early to draw firm conclusions on the extent to which such forums may contribute to enhancing compliance within the anti-discrimination field. However, reference must be made to the suggestion that, at the current moment, ‘there is still ample evidence of local diversity’,¹⁹¹ and that the **differences between domestic institutional structures and socio-political contexts** challenge any idea that best practices identified for one Member State are readily useful and/or transferable to the others.

Indeed, the extent to which the transposition, as well as the application and enforcement, of the anti-discrimination directives are realised in practice is highly contingent both on the extent to which the ideas embodied by the directives resonate within the domestic political and societal environment, and the degree to which the institutional arrangements

¹⁸⁷ European Network of Legal Experts in the Non-Discrimination Field, European Anti-Discrimination Law Review, No. 1, 2005, p. 15, 38.

¹⁸⁸ Bell, M., The Implementation of European Anti-Discrimination Directives: Converging Towards a Common Model?, The Political Quarterly, 2008, p. 41.

¹⁸⁹ Commission of the European Communities, Communication From the Commission to the Council and the European Parliament, the European Economic and Social Committee and the Committee of the Regions, Non-Discrimination and Equal Opportunities: A Renewed Commitment, COM(2008) 420 final, 2008, p. 9.

¹⁹⁰ See the discussion regarding the MLP earlier in this chapter.

¹⁹¹ Bell, M., The Implementation of European Anti-Discrimination Directives: Converging Towards a Common Model?, The Political Quarterly, 2008, p. 43.

are ‘fit for purpose’.¹⁹² Thus, even if the clearly **iterative nature** of the learning process and the (attempted) inclusion of a broad range of stakeholders may stimulate policy learning – and, possibly, policy convergence – across the board, in order for learning to be truly effective, there must be recognition of cross-national differences and an understanding that an ‘adaptive transfer’ of ‘good practices’ in a policy mix that best suits each Member State is the most that we can expect to achieve.¹⁹³

The extent to which the Commission and other affected actors are aware of this important point is not entirely clear from the official documents that were consulted for this analysis. However, the suggestion that some of the norms found within the directives themselves are of a very **open nature**, and that space was deliberately left for **domestic experimentalism** (for example, with regard to the structure of the equality bodies),¹⁹⁴ may suggest that there was a recognition that a ‘one-size-fits-all’ approach would not be sufficient to make the Member States comply.

4. The need for political determination – revisited

Whatever the case, problems related to compliance have arisen in the past and continue to exist today.¹⁹⁵ Even if these problems may be partly explained by mismatches between the philosophy and/or practical obligations embodied in the EU legislation, and the prevalence of domestic philosophies, laws and/or administrative structures,¹⁹⁶ the lack of a **genuine commitment** on the part of the Member States to effectively implement the EU’s anti-discrimination legislation is also an important causal factor. In fact, it has been noted that the weak political determination in many Member States, and a widespread ignorance of the need for anti-discrimination legislation, have resulted in failures in transposition or ‘grudgingly transposed weak legislation, destined to be under-promoted by national governments’.¹⁹⁷

Therefore, even if compliance is eventually legally enforceable through the Commission’s power to commence infringement proceedings against non-compliant

¹⁹² It is suggested that, in fact, the implementation of the directives required significant changes in most Member States, which suggests that there was in practice more often than not a ‘misfit’, rather than a ‘good fit’, between EU and national policies within this field. See Geddes, A. and Guiraudon, V., Britain, France, and EU anti-discrimination policy: the emergence of an EU policy paradigm, *West European Politics*, 2004.

¹⁹³ See also our discussion on the need for a ‘context sensitive, lesson drawing approach’ earlier in this chapter.

¹⁹⁴ Bell, M., *The Implementation of European Anti-Discrimination Directives: Converging Towards a Common Model?*, *The Political Quarterly*, 2008, p. 43.

¹⁹⁵ As a matter of fact, the Commission has started infringement proceedings against a number of Member States, either because they had not notified the Commission of transposition of the anti-discrimination directives, or because they had not transposed the directives in a timely manner and/or correctly.

¹⁹⁶ See for example Bell, M., *The Implementation of European Anti-Discrimination Directives: Converging Towards a Common Model?*, *The Political Quarterly*, 2008; Commission of the European Communities, Report from the Commission to the Council and the European Parliament, *The Application of Directive 2000/78/EC of 27 November establishing a general framework for equal treatment in employment and occupation*, COM (2008) 225 final 2, 2008.

¹⁹⁷ European Network of Legal Experts in the Non-Discrimination Field, *European Anti-Discrimination Law Review*, No. 1, 2005, p. 32.

Member States, and individuals may in certain cases bring national governments before the court where cases of non-compliance are concerned, it is equally true that sustained political determination at the highest levels would help to facilitate the implementation of legislation. This is perhaps particularly so when such political determination represents a challenge to national norms and institutions, and prevents resort to instruments that are should be used only as a last resort.

4.7 Lessons confirmed – Part II: the Stability and Growth Pact and the ‘Brussels I Regulation’¹⁹⁸

The above analysis on the functioning of the evaluation mechanisms within the anti-discrimination field identifies many important factors that have also arisen within the case study presented in the main part of this chapter. It is very interesting to see that, whatever policy area we examine, similar issues seem to figure prominently as either problematic or successful. For example, positive experiences with the Mutual Learning Programme have stimulated several Directorates-General of the Commission to implement similar types of peer review mechanisms in their fields of competence.¹⁹⁹ Peer review exercises have now become common evaluation tools aimed at stimulating compliance, even in rule-based integration frameworks, where it might initially appear that there is little room for peer review activities.

In addition to this being the case within the anti-discrimination field, the *Stability and Growth Pact* (SGP) constitutes another prominent example of a binding policy framework where ‘peer support and peer pressure are [nevertheless] essential to ensure a proper and effective implementation of the rules’.²⁰⁰ The SGP, which lays down fiscal rules for the Member States to comply with, consists of a preventive arm focusing on multilateral surveillance and the avoidance of excessive deficits, and a dissuasive arm, where the Member States in the Council are expected to sanction excessive deficits when they arise.

However, the SGP is also a very clear example of the axiom that, in order for the theoretical potential of a peer review mechanism to materialise, a **firm political commitment** on the part of the Member States to make the mechanism work is necessary. The fact that the Member States simply did not want to move on to the sanctioning phase recommended by the Commission, when France and Germany continued to experience excessive deficits in 2003, demonstrates that, without such political commitment, the system breaks down, loses its credibility and remains ‘toothless’.

¹⁹⁸ Council Regulation (EC) No 44/2001 on jurisdiction, recognition and enforcement of judgments in civil and commercial matters.

¹⁹⁹ This is not to neglect the fact that, in several policy fields, some forms of peer review activities have existed for a longer period. However, it must be emphasized that the MLP is often referred to as a positive reference point for other policy fields, stimulating them either to introduce similar instruments, or to adjust their existing peer evaluation framework.

²⁰⁰ Bertoldi, M., Peer Review, Peer Support and Peer Pressure: The European Union Experience, in OECD, 2008.

Although the failure of the system to effectively enforce compliance in these cases may have constituted the most visible trigger for the reforms of the SGP in 2005, a number of other factors that also precluded the effectiveness of the peer review mechanism clearly loomed in the background. For example, the practice of **self-reporting** by Member States appeared to have resulted in instances of misreporting, which could, in some cases, also (partly) be traced back to a **lack of independence of the national statistical agencies** from domestic governments. Moreover, a lack of ownership of the peer review process at the national level was problematic, with the **absence of national watchdog bodies** one of the greatest lacunae.

Other factors complicating the evaluation of Member States' performances related to the existence of a **small number of fiscal indicators**, and **outcome-based, rather than input-based, assessments** of compliance. Finally, the **absence of appropriate incentives** during periods of economic upturn so as to avoid falling into the excessive deficit procedure at a later stage, and a complete **disregard of country-specific circumstances**, were noted as points of concern.²⁰¹

Even if the 2005 reform of the Pact seems to go some way to addressing the problematic issues, there still seems room for improvement of, in particular, the preventive – multilateral surveillance – arm of the SGP.²⁰² As can be seen,, many of the issues identified as problematic regarding this area have also had a similar effect on the MLP.

We will now examine the evaluation process regarding the application of Council Regulation (EC) No 44/2001 on jurisdiction, recognition and enforcement of judgments in civil and commercial matters, or the *Brussels I Regulation*,²⁰³ which constitutes another interesting case in relation to compliance by the Member States with binding rules. The reporting method employed for this Regulation reminds us of the **'expertocratic' reporting procedure** that is used for evaluations within the anti-discrimination field: independent experts acting as 'coordinators' of the evaluation exercise drew up **standardized questionnaires** that were circulated amongst national experts, who collaborated as a network of correspondents. On the basis of these questionnaires, the experts consulted electronic databases and conducted **interviews with the relevant stakeholders** in their respective countries – given the problematic **lack of statistical data** in many Member States, interviewing the relevant stakeholders was considered to be the most reliable method of data collection.

Additionally, foreign experts were contacted by the project coordinators, in order to collect as much information as possible. Part of the process also involved a **seminar**, where the results of the empirical research and possible improvements were discussed by the national experts and coordinators. The final report was completed following

²⁰¹ Annett, A., Decressin, J. and Deppler, M., Reforming Stability and Growth Pact, IMF Discussion Paper, PDP/05/02, 2005, p. 115.

²⁰² Commission of the European Communities, Public Finances in EMU – 2008, Brussels, European Commission, 2008.

²⁰³ Council Regulation (EC) No 44/2001 on jurisdiction, recognition and enforcement of judgments in civil and commercial matters.

discussions involving the project coordinators, the national reporters and the Commission.²⁰⁴ As noted earlier, this expert-centred way of reporting is more likely to produce the independent information necessary for the Commission to effectively function as ‘guardian of the Treaties’.

However, in this case we also see a certain degree of reliance on **ranking**, **‘naming and shaming’** and **mutual learning** as mechanisms for stimulating compliance. The final evaluation report published by the project coordinators clearly presents a **comparative study** on the performances of the Member States, as a result of which non-compliers may feel pressured to step up their efforts to comply with the Regulation, perhaps because they would also want to avoid the imposition of ‘hard sanctions’ that exist in the background, as coercive mechanisms enforcing compliance.

However, the report is also clearly intended to be a source of learning: as it identifies in detail compliance problems that Member States have dealt with and also expressly stresses the ‘best practices’ in the Member States related to the application of the Regulation, there are ample opportunities for policy learning by civil servants and other affected stakeholders.²⁰⁵ The particular elements addressed are also similar to the way in which the anti-discrimination reports are intended to stimulate learning and compliance.

Although the list of interesting cases to be examined within the context of this chapter is, of course, non-exhaustive, the additional analyses above confirm our recommendations on how to improve the functioning of evaluation mechanisms, as presented in section 4.5. Indeed, very often the same factors hinder the complete realisation of the theoretical potential of evaluation mechanisms. In many cases, structures that have apparently proven their usefulness for evaluating the performance of Member States are used. Together with the analysis concerning the functioning of evaluation mechanisms in the third pillar, which is presented in chapter 5, the conclusions arrived at in this chapter may, therefore, provide a useful basis on which to contemplate scenarios for improving the current evaluation mechanisms in the third pillar.

²⁰⁴ Hess, B., Pfeiffer, T. and Schlosser, P., Report on the Application of Regulation Brussels I in the Member States, Study JLS/C4/2005/03, 2007.

²⁰⁵ Hess, B., Pfeiffer, T. and Schlosser, P., Report on the Application of Regulation Brussels I in the Member States, Study JLS/C4/2005/03, 2007.

5. Monitoring and evaluation mechanisms at work in the third pillar

In paragraph 3.3, the evaluation mechanisms applicable to the third pillar were presented. The picture that emerges from that presentation is that in the third pillar, there has also been some experience with evaluation. How should the existing mechanisms be characterised? We will characterise the current applicable evaluation mechanisms as *implementation-biased*. This analysis and the recommendations for improvement of the existing mechanisms are found below. In simple words, it comes down to *what* is evaluated, *how* by *whom* and what are the *consequences* to be attached to the findings.²⁰⁶

The five differentiated mechanisms as presented in paragraph 3.3 are discussed together below, in order to allow for an overall and coherent analysis of the mechanisms. However, the remarks and observations relate primarily to the peer evaluation mechanism set up by Joint Action 97/827 of 5 December 1997 and the Schengen evaluation mechanism. Where other mechanisms are referred to, this will be specifically indicated.

5.1 Analysis of the existing evaluation mechanisms

The instruments evaluated are binding legal instruments

Generally speaking, the instruments subject to evaluation and monitoring are legal instruments containing binding norms. Occasionally, action plans and other policy documents are evaluated. On its own initiative, the European judicial network has undertaken an evaluation itself.²⁰⁷

Regarding the evaluation by the Commission of the implementation of legislation by the Member States, the following can be noted. In addition to the binding character of the norms, it is also required that the Commission will draw up a report, some time after the implementation period has lapsed. Many Framework Decisions contain the obligation that, after a certain period of time, a report shall be drafted by the Commission. This report relates to formal implementation only.²⁰⁸ Evaluations as such are binding for Member States, because that is stipulated in the legal instrument. However, no sanctions are provided if Member States do not submit the information on how they transposed their obligations into national law. There is no mechanism to require compliance with the outcome.

Compliance with implementation

Member States do not always comply in a timely manner (or at all) with the obligation to communicate national transposition legislation regarding Framework Decisions. On occasions, there has been no action even after more than two years following the

²⁰⁶ Serge de Biolley and Anne Weyembergh, L'évaluation dans le cadre du troisième pilier du traité sur l'Union européenne, in: Weyembergh, A. and S. de Biolley, Comment évaluer le droit pénal européen?, Bruxelles 2006, p. 76.

²⁰⁷ See European Judicial Network, Replies to the questionnaire on the evaluation of the tools for judicial cooperation in criminal matters, Council Document 5684/09, 26 January 2009.

²⁰⁸ For example, Article 34 Framework Decision 2002/584, Article 10 Framework Decision 2002/629 and Article 14 Framework Decision 2003/577.

implementation deadline.²⁰⁹ There is no Commission evaluation report that states that all Member States had responded to it in time. Often, no Member State, or perhaps just a handful, reported on time. Occasionally, the Commission delayed its report, because it had insufficient Member State responses to address.²¹⁰ However, it must be noted that not communicating does not necessarily mean that there was no transposition. Some Member States did implement without communicating it.

The Commission has formulated two indicators for compliance. The first is the communication of the implementing legislation to the Commission by the Member States. The second is the correct implementation by the Member States.²¹¹ The Commission found that Member States failed to meet their obligations on both indicators: Member States do not communicate whether they have implemented or not, and do not correctly implement their obligations.²¹²

Member States do seem more eager to communicate their legislation on fighting terrorism. As described in 3.3, a specific regime of evaluation and monitoring applies in that area. One may conclude that the political urgency of a piece of legislation contributes to compliance.

The methodology used for peer evaluation

The methodology generally chosen for the peer evaluation comprises the following components: a questionnaire, on the basis of which a peer review takes place.²¹³ The peer review consists of a visit to the Member State to be evaluated, a report drawn up by the evaluation team and subsequent discussion of the report. The visits by a team of experts from the Member States and the Institutions are announced.²¹⁴ Joint Action 97/827 merely states that the General Secretariat of the Council shall draw up a questionnaire. This is a task given to the Committee of Permanent Representatives of the Member States in Joint Action 98/429 and Decision 2002/996 to CATS.

No criteria have been formulated as to what must be evaluated, how the effects of legislation are to be measured and according to which methodology a questionnaire should be drafted, whether the evaluation deals with an evaluation plan (why are we

²⁰⁹ Communication from the Commission to the Council and the European Parliament, Report on Implementation of The Hague Programme for 2007, Brussels, 2.7.2008, COM (2008) 373 final, p. 14.

²¹⁰ See, for example, COM (2001) 771.

²¹¹ It seems that the Commission relies for this on letters from citizens and questions raised in the European Parliament, or on the basis of reports from the Commission itself.

²¹² Mededeling van de Commissie aan de Raad en het Europees Parlement, Verslag over de uitvoering van het Haags programma- 2005, Brussel, 28.6.2006, COM (2006) 333definitief.

²¹³ This is provided for in Joint Action 97/827.

²¹⁴ The Commission gave the impression that it would favour unannounced visits. However, in the final version of its proposal establishing an evaluation mechanism of the Schengen acquis, they are expressly not provided for. See Voorstel voor een Besluit van de Raad betreffende de instelling van een evaluatiemechanisme voor toezicht op de toepassing van het Schengen acquis, Brussel, 4.3.2009 COM (2009)105 definitief, 2009/0032 (CNS). Proposal for a Council Regulation on the establishment of an evaluation mechanism to verify the application of the Schengen acquis, 4.3.2009 COM (2009) 102 final, 2009/0033 (CNS).

doing this?), a process evaluation (did we evaluate well?), or whether the evaluation is performed as a distraction from politically sensitive issues. This runs the risk that the questionnaires may differ from evaluation to evaluation. On another level, questionnaires also carry the risk of uneven comparisons: both the evaluating team and the respondents differ from State to State.

The Member States are the source of information for the evaluations. The Commission is currently completely dependent on the Member States for its information.²¹⁵ The Commission recognizes its dependence on national reports: '[t]he value of this report depends therefore largely on the quality of the national information received by the Commission.'²¹⁶ There is a risk of bias if the answers to the questionnaire and receiving the evaluation team are carried out by the same civil servant(s) who are responsible for implementation. Evaluation is often based on incomplete information (not all Member States) and may therefore lead to uneven comparisons. It raises the question as to whether it would be possible that information on the Member State's practice will reach the evaluators via independent or other sources.

The methodology used under the Schengen mechanism

Similarly, an evaluation mechanism has been drafted (both as a Decision and a Regulation) regarding the application of the Schengen acquis.²¹⁷ It is provided that evaluations take place on the basis of questionnaires and that announced on-site visits may take place.²¹⁸ This involves the whole Schengen acquis in criminal law. The on-site visit takes place after receiving the reply to the questionnaire.²¹⁹ Also, in this area, it is often the case that only a limited number of Member States reply to the questionnaire.²²⁰ It is also noted that the new drafts pay very little attention to methodology.

The methodology used in the Commission reports

²¹⁵ Serge de Biolley and Anne Weyembergh, L'évaluation dans le cadre du troisième pilier du traité sur l'Union européenne, in: Weyembergh, A. and S. de Biolley, Comment évaluer le droit pénal européen?, Bruxelles 2006, p. 84.

²¹⁶ For example, COM (2001) 771.

²¹⁷ Voorstel voor een Besluit van de Raad betreffende de instelling van een evaluatiemechanisme voor toezicht op de toepassing van het Schengenacquis, Brussel, 4.3.2009 COM (2009)105 definitief, 2009/0032 (CNS). Proposal for a Council Regulation on the establishment of an evaluation mechanism to verify the application of the Schengen acquis, 4.3.2009 COM (2009) 102 final, 2009/0033 (CNS).

²¹⁸ Questionnaires and on the spot evaluation are provided for in Joint Action 97/827. The same goes for the European Arrest Warrant. See Paola Fiore, The Fourth Round on Mutual Evaluation Carried out by the Council of the European Union of the Practical Application of the European Arrest Warrant, 8 ERA Forum 2007, p. 225 - 232.

²¹⁹ See Article 6 Joint Action 97/827, which has the effect that an on-site visit can only take place after the Member State has submitted its reply.

²²⁰ See, for example, Second report on the state of implementation by Member States and EU bodies of the action-Oriented Paper on Improving Cooperation on Organized Crime, Corruption, Illegal Immigration and Counter-terrorism between the EU, the Western Balkan and relevant ENP countries, Brussels 22 May 2008, 8827/2/08.

In the evaluation reports on Framework Decisions, the Commission regularly refers to its report on the implementation of Community waste legislation, as the document in which it stated its approach towards evaluation of legislation. What is apparent from that report is that the Commission is not satisfied with the way evaluations are conducted. Although its main criticism relates to the scope of the evaluations, it also criticizes the method, the timing and the self-reporting by the Member States. At the end of the report, the Commission gives the following recommendation regarding future reports and questionnaires:

‘At present the questionnaires and thus the report on the implementation of waste legislation is a mixture between the legal transposition and the practical implementation of Community legislation. However, this approach should be reconsidered. Indeed, it does not seem to be opportune to establish reports on the application of Community waste legislation every three years, which, to a large extent, informs on the legal transposition of Community waste directives into national law. Further, the legal conformity of the national law should be checked once after having transposed Community law or again after the amendment of the national law; in contrast, the reports should much more focus on the experience made by the application in practice. For this reason the questionnaires of Commission Decision 94/741/EC of 24 October 1994 and 97/622/EC of 27 May 1997 might have to be adapted. In addition, Annex VI of Directive 91/692/EEC should be adapted in order to comply with current and partly amended Community waste legislation.

The reports on the implementation of Community law present an important tool for the Commission's task to act as the guardian of the EC Treaty. However, it has to be noticed that the reports are mainly based on contributions from Member States themselves. This obviously limits the possibility to identify omissions of applications or weaknesses and lacunes [*sic*] of existing Community waste legislation.²²¹

While the Commission at times indicates that it is unsatisfied with the current mechanisms of evaluation and formulates ideas for improvement, there is no consistency in its approach. Given the rather discretionary instruction in the Framework Decisions - ‘a written report by the Commission’ - there is nothing that would prevent the Commission from choosing another methodology for evaluation.

The evaluation team in the peer evaluation mechanism

The evaluation teams consist of civil servants both from other Member States and the Commission, and vary in their composition for each individual Member State being

²²¹ Report from the Commission to the Council and the European Parliament on the implementation of Community waste legislation Directive 75/442/EEC on waste, Directive 91/689/EEC on hazardous waste, Directive 75/439/EEC on waste oils and Directive 86/278/EEC on sewage sludge for the period (1995 - 1997) /* COM/99/0752 final */ , p. 92.

evaluated. Coordination of the evaluation is often the responsibility of the (General Secretariat of the) Council. ‘Experts with substantial practical knowledge of the EAW were nominated by Member States²²² pursuant to a written request to delegations made by the Chairman of the MDG.’²²³ A specific procedure is applied in Decision 2002/996 regarding the evaluation of measures against terrorism. The Member States suggest experts to be listed on a list of experts, from which the Presidency selects appropriate persons for each evaluation visit.

The specific nature of the personnel to be used for each team does not put pressure on a Member State being evaluated. The peer expert of today will itself be evaluated tomorrow.²²⁴ Andersen characterised the non-binding peer review as ‘a relatively “sovereignty-responsive” method compared to the general infringement procedure.’²²⁵ She further argues that, although the peer review mechanism may detect and highlight compliance problems, it is primarily concerned with contested norms. When non-compliance is rooted in a dispute over the law, the Commission has no authoritative power of interpretation.

It should be noted that the evaluation team must have the expertise to use the methodology chosen for evaluation, as well as to apply the specific evaluation criteria. Since the current mechanisms use legal methods and legal criteria, an evaluation team comprised of lawyers meets the requirements. The teams are qualified to apply the methodology used. However, it is submitted that the methodology should go beyond just the law in the books.

Criteria for evaluation in the Commission report

The Commission has formulated general criteria for evaluation as follows:

- ‘1. form and methods of implementation must be chosen in a manner which ensures that the directive functions effectively with account being taken of its aims;
2. each Member State is obliged to implement directives in a manner which satisfies the requirements of clarity and legal certainty and thus transpose the provisions of the directive into national provisions, which have binding force;
3. transposition need not necessarily require enactment in precisely the same words in an express legal provision; thus a general legal context (such as already

²²² One could perhaps argue, therefore, that the emphasis is more on the expert being a representative of a specific Member State, than on his or her status as an expert.

²²³ Evaluation report on the fourth round of mutual evaluations practical application of the European Arrest Warrant and corresponding surrender proceedings between Member States. Report on Spain, Brussels, 6 June 2007, 5085/2/07, p. 3. Idem concerning The Netherlands, 2 December 2008, 15370/1/08.

²²⁴ De Schutter, 2008, p. 17. The Commission suggested outsourcing. See Discussion Paper, Meeting of Experts on evaluation in the JHA area, Brussels – Wednesday 18 February 2009. However, it is clear that it wants to maintain control over what other entities might do in their investigations.

²²⁵ Stine Andersen, *Sovereignty and the emergence of non-binding peer review within the EU*, on file with the authors, p. 15.

existing measures, which are appropriate) may be sufficient, as long as the full application of the directive is assured in a sufficiently clear and precise manner;
4. directives must be implemented within the period prescribed therein.²²⁶

It is striking that evaluation criteria have not been formulated for the other mechanisms.

The scope of evaluation: implementation

The scope of evaluation in the existing mechanisms is limited by two factors:

- 1) the legal instrument;
- 2) implementation of its obligations within the internal law of the Member States.

The need, impact or efficiency of the Union's legislation is not evaluated.²²⁷ Neither is the (role of the) Commission itself ever evaluated.²²⁸ The evaluations look almost exclusively at Member States' obligations. The scope of the evaluations further depends on the legal instrument being evaluated. Most evaluation reports on the Framework Decisions focus on the question as to whether the obligations from the legal instrument have been implemented. Only the Report on the EAW stipulated specific criteria of evaluation, focussing on the judicial nature, efficiency and rapidity of the EAW.²²⁹

The scope of evaluation: the rule of law

In the context of enlargement, a mechanism for collective evaluation of applicant countries had already been established in 1998.²³⁰ In order to assess possible problem areas, 'all relevant material to the enactment, application and effective implementation by the candidate countries (...) shall be made available,' in accordance with Article 3 Joint Action 98/429. Additionally, the group of evaluation experts may obtain information from other sources, including Council of Europe reports. For example, the state of the

²²⁶ These criteria were developed in the context of Directives, appearing for the first time regarding Framework Decisions in COM (2001) 771. They have been used ever since. See Report from the Commission on the implementation since 2005 of the Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between the Member States, p. 3.

²²⁷ However, Decision 2006/581 allows for consultation of experts in identifying the policy needs for data on crime and criminal justice at the EU level. See further Weyembergh and Santamaria, Introduction, in: A. Weyembergh and V. Santamaria, *The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings*, Bruxelles 2009, p. 15-16.

²²⁸ See S. de Biolley and A. Weyembergh, *L'évaluation dans le cadre du troisième pilier du traité sur l'Union européenne*, in: *Comment évaluer*, p. 79-80; Olivier De Schutter, *The Role of Evaluation in Experimentalist Governance: Learning by Monitoring in the Establishment of the Area of Freedom, Security and Justice*, Faculté de Droit de L'Université de Louvain, November 2008, p. 10 – 11.

²²⁹ See SEC (2005) 267, p. 4.

²³⁰ Joint Action 98/429 of 29 June 1998 adopted by the Council on the basis of Article K.3 of the Treaty on European Union, establishing a mechanism for collective evaluation of the enactment, application and effective implementation by the applicant countries of the acquis of the European Union in the field of Justice and Home Affairs, OJ L 191, 7.7.1998, p. 8–9. See also Council regulation 622/98 of 16 March 1998 on assistance to the applicant States in the framework of the pre-accession strategy, and in particular on the establishment of Accession Partnerships, OJ 1998, L 85/1.

rule of law in the then candidate States of Romania and Bulgaria was recognized as a risk.²³¹

Candidate Member States are very eager to comply with the demands of EU law. A high degree of compliance is noted. However, after accession, ‘compliance fatigue’ may emerge.²³²

The standards upon which compliance with the rule of law is tested have been set by the JHA Council Conclusions of 28 May 1998. These mention the following as the main elements of the rule of law:

- independence of the judiciary;
- effective access to justice;
- respect for judicial decisions, and
- an objective system of public prosecutions.

Outside the field of enlargement, it has only been recently that attention has been focussed on the rule of law, this coming with the establishment of the Justice Forum. This Forum tries to obtain a picture of whether the European area of justice is operating efficiently and if those using and working in it are satisfied.²³³ It is vital that the scope of evaluation is broadened from an assessment of the effect of legislation on national systems, to one of the effects for the Union as a whole.²³⁴ In Joint Action 97/827, the objective of the evaluation is Union legislation and ‘other international acts and instruments in criminal matters.’²³⁵ There is a link with the effectiveness of the system. It is submitted that there is some flexibility to evaluate in this field.

Respect for the rule of law received further support when, in March 2008, the Netherlands emphasized the relevance of common standards in procedural law, processes and procedures and the necessity of monitoring the rule of law.²³⁶ It was well received by the Commission. Commissioner Barrot stated recently that it is necessary to ‘mettre la personne au centre de la construction de l’espace JLS’.²³⁷ Europe must protect the

²³¹ See Commission Decision of 13/XII/2006 establishing a mechanism for cooperation and verification of progress in Romania to address specific benchmarks in the areas of judicial reform, and the fight against corruption, Brussels 13/XII/2006, C (2006) 6569 final; Commission Decision of 13/XII/2006 establishing a mechanism for cooperation and verification of progress in Bulgaria to address specific benchmarks in the areas of judicial reform, and the fight against corruption, Brussels 13/XII/2006, C (2006) 6570 final.

²³² See P. de Hert and E. Pitto, Evaluation in the context of enlargement, in Weyembergh, A and S. de Biolley, *Comment évaluer le droit pénal européen?*, Bruxelles 2006, p. 163-195; K. Šugman, Critical analysis of evaluation process in assessing Candidate Countries, *idem* p. 197-207; R. Genson and W. van de Rijt, *L’évaluation de Schengen dans le cadre de l’élargissement*, *idem*, p. 209-218.

²³³ See Communication from the Commission on the creation of a Forum for discussing EU justice policies and practice /* COM/2008/0038 final, paragraphs 2, 9, 11 and 16.

²³⁴ See further Weyembergh and Santamaria, Introduction, in: A. Weyembergh and V. Santamaria, *The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings*, Bruxelles 2009, p. 10.

²³⁵ Article 1 Joint Action 97/827.

²³⁶ Report by the Future Group (Justice), Brussels, 22 July 2008 (23.07), 11960/08, p.9.

²³⁷ Colloque ‘Preparer le Programme de Stockholm’ Collège d’Europe, Bruges, le 4 mars 2009.

individual and must also include his protection in evaluation mechanisms. Compliance with general principles of law, the rule of law and human rights obligations has thus far never been within the scope of a monitoring mechanism. This is different at the level of the United Nations.²³⁸

It is surprising to see that, in the political discourse (followed by legislation and the Court's case law) of the Member States and the Commission, there is a complete absence of understanding that that mutual trust in practice is not unconditional.²³⁹ This does not demonstrate political realism.²⁴⁰ A relationship between the level of harmonization of procedural law and procedural safeguards, and mutual trust, has been identified.²⁴¹ The minimum that seems to be accepted is that 'mutual trust must be further enhanced.'²⁴²

The Dutch Minister of Justice identified evaluation and monitoring systems as one of the ways in which such mutual trust can be guaranteed.²⁴³ It is submitted that improving the application of the rule of law in all Member States will strengthen mutual trust and, as a consequence, facilitate mutual recognition.²⁴⁴

Frequency and timing of evaluation exercises

Joint Action 97/827 provides that at least five Member States per year will be evaluated, with the first year of evaluations beginning no later than three months after the entry into force of the Joint Action.²⁴⁵ It appears that both the peer evaluations, as well as the evaluations conducted by the Commission, take place at a time when there is little more

²³⁸ The United Nations also took initiatives regarding the respect for the rule of law. See the Vera Institute of Justice, Rule of Law Indicator Instruments: A Literature Review, A Report to the Steering Committee of the United Nations Rule of Law Indicators Project, November 2008. See further two contributions in Kauko Aromaa and Slawomir Redo (eds.), for the Rule of law: Criminal Justice Teaching and Training @cross the World, Helsinki - Seoul 2008: Jean-Marc Coicaud, The United Nations in the International Security and Criminal Justice Politics: Making the Rule of Law Work, p. 45-52; Yvon Dandurand, Curt T. Griffiths and Vivienne Chin, Training and Effective Support of Comprehensive Justice and Security Reforms: Outcomes to Date and Lessons Learned, p. 53-67.

²³⁹ G. Vernimmen-Van Tiggelen and L. Surano, Analysis of the future of mutual recognition in criminal matters in the European Union, Université Libre de Bruxelles 2008, p. 20.

²⁴⁰ In the context of the European arrest warrant, for instance, many Member States still take the view that they are generally entitled to refuse the surrender of their nationals. See Conclusion of AG Bot of 24 March 2009, Dominic Wolzenburg, C-123/08, par. 120-125.

²⁴¹ G. Vernimmen-Van Tiggelen and L. Surano, Analysis of the future of mutual recognition in criminal matters in the European Union, Université Libre de Bruxelles 2008, p. 10.

²⁴² Report on the Informal Meeting of the Ministers for Justice and Interior on 15 and 16 January 2009.

²⁴³ Speech at Maastricht University on 2 June 2009.

²⁴⁴ See Weyembergh and Santamaria, Introduction, in: A. Weyembergh and V. Santamaria, The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings, Bruxelles 2009, p. 9; I. Jegouzo, La création d'un mécanisme d'évaluation mutuelle de la justice, corollaire de la reconnaissance mutuelle, p. 152.

²⁴⁵ The Council considered that evaluations should initially be conducted every five years, Presidency, Evaluation of EU-policies on Freedom, Security and Justice - Council Conclusions, Brussels, 19 June 2007, 10893/1/07.

to evaluate than the implementation of legislation.²⁴⁶ There is hardly any practice or case law to be taken into consideration. The timing thus contributes to the focus of the evaluation being on implementation.

When the Framework Decision on the European Arrest Warrant was evaluated very soon after entering into force, the reports offered more information than on implementation alone. However, this cannot be explained by its timing, but rather by the fact that this is the only instrument on which all Member States have implemented national legislation. In relation to those legal instruments for which implementation is incomplete, the determination of the most appropriate timing for evaluation is difficult. This also raises questions about the initial necessity of the legal instrument.

Specifically with regard to peer evaluations, apart from the frequency, the duration of the evaluation exercise is also an issue. In the context of the mutual evaluation of the exchange of information and intelligence between Europol and the Member States, visits were paid to the Member States over a period of four years.²⁴⁷ The length of this period makes it difficult to maintain an equal standard of comparison between the Member States.

Examples of ‘evaluation fatigue’ have also been identified. This relates to the number of evaluations, their timing, duration and frequency. It is most likely that this evaluation fatigue has a negative effect on implementation, response to questionnaires and compliance with the findings of evaluation exercises.²⁴⁸

Findings

As has been noted several times in this report, most reports concentrated on implementation issues. The legal instrument itself is not evaluated, but rather the way Member States have implemented it or use it in practice. However, occasionally,²⁴⁹ some of the findings may be relevant for the appropriate functioning of the legal instrument at hand. The result is then an inventory of all kinds of practical problems and some recommendations on how to solve them.²⁵⁰ Recommendations may be made for the Member State evaluated, other Member States and/or the European Union. It does not, however, necessarily lead to a further clarification of common principles, such as the

²⁴⁶ Marie-Aude Beernaert, De quelques considérations profanes sur l'évaluation de l'impact, au niveau national, des lois pénales internes, in: Weyembergh, A and S. de Biolley, Comment évaluer le droit pénal européen?, Bruxelles 2006, p. 30.

²⁴⁷ Final report on the evaluation visits of all 27 Member States, Brussels, 23 November 2007, 13321/3/07 REV3.

²⁴⁸ To determine exactly how influential this is, non-legal research should also be conducted.

²⁴⁹ Whenever this has occurred, it has concerned peer evaluations, not Commission reports.

²⁵⁰ See, for example, Evaluation report on the fourth round of mutual evaluations the practical application of the European Arrest Warrant and corresponding surrender procedures between Member States” Report on Spain, Brussels, 6 June 2007, 5085/2/07, REV2.

principle of mutual recognition. Evaluations have been characterized as a compensation for the lacking infringement procedures.²⁵¹

It has been reported that the introduction of the European Arrest Warrant mechanism is a great success. However, the question is whether it is possible to reach this conclusion in the absence of a standard of comparison. A comparison between the situation with the new EU legislation and the previous situation is absent in current evaluation mechanisms. The European Arrest Warrant scheme replaced the existing extradition scheme and was, for a long time, the only applicable instrument of mutual recognition. It should be the case that not only the successes of the system are highlighted in the report.

The peer evaluation team prepares a draft report, which includes recommendations. Before the report is adopted, the Member State in question may make observations, which are then discussed and possibly included in the report. According to Joint Action 97/827, Member State's comments can lead to an amendment of the report if the team deems that relevant. Referring to the European Arrest Warrant, Fiore states: 'If an agreement is not reached between the Member State and the expert team, the disputed remarks are annexed to the report.'²⁵² From a methodological point of view, the practice that both findings and recommendations are negotiable for the Member States being evaluated is highly problematic. It infringes upon the independence and reliability of the mechanism.

Accessibility of the findings

According to Article 8 and 9 of the Joint Action 97/827, which establishes the regime for peer evaluation, the report will not be shared with other Member States and will remain confidential, unless the Member State concerned wishes to publish it. Most reports on the EAW are available at the website of the Council. Article 9 of the Joint Action 97/827 obliges the Presidency to inform the European Parliament each year of the implementation of the evaluation mechanism. We have been unable to trace what the European Parliament does with this information.²⁵³

If collective learning is to be regarded as one of the goals of evaluation and monitoring, then publication of the results is necessary. If Member States do not have access to reports concerning other Member States, they cannot learn from those practices and examples. Joint Action 98/427 provides for statements concerning good practice. This may lead to collective learning.

²⁵¹ Olivier De Schutter, *The Role of Evaluation in Experimentalist Governance: Learning by Monitoring in the Establishment of the Area of Freedom, Security and Justice*, Faculté de Droit de L'Université de Louvain, November 2008, p. 9.

²⁵² See Paola Fiore, *The Fourth Round on Mutual Evaluation Carried out by the Council of the European Union of the Practical Application of the European Arrest Warrant*, 8 ERA Forum 2007, p. 231.

²⁵³ The European Parliament recommends an effective role for itself in a new evaluation mechanism. See *Verslag met een aanbeveling van het Europees Parlement aan de Raad betreffende de ontwikkeling van een strafrechtsgebied in de EU (2009/2012 (INI))*, 8.4.2009, A6-0262/2009.

The reports drafted by the Commission on implementation, as well as the ‘scoreboards’, are publicly available.

(no) follow up

Generally speaking, there is no follow up to evaluation exercises.²⁵⁴ Even the most important legal instrument on peer evaluation, Joint Action 97/827, does not mention a follow up of the evaluation results. It aims to allow for the ‘evaluat[ion] on a basis of equality and mutual confidence the implementation.’ Article 8 stipulates that ‘the Council, may, where it sees fit, address any recommendations to the Member State concerned and may invite it to report back to the council on the progress it has made by a deadline to be set by the Council.’²⁵⁵ It is both necessary that a follow up to the results of the evaluation will be undertaken, as well as that this will be institutionalised.²⁵⁶

Sanctions for non-compliance with evaluation findings are not provided for at all. Although Article 35, paragraph 7 TEU allows Member States to do so, no Member State has ever brought a case before the Court on the interpretation of a third pillar instrument against another Member State.²⁵⁷

Similarly, regarding obtaining information for Commission reports, the only thing the Commission can do is to remind Member States and establish new deadlines. See, for example, the ‘powerless’ request of the Commission in COM (2004) 230: ‘In view of the foregoing, the Commission invites the Member States to ensure a rapid and complete transposal of the Framework decision and to inform it of this immediately, and no later than the 1st of September 2004, providing a description of the measures taken with the text of the statutory or administration provision in force in support.’²⁵⁸

Although the various stakeholders demonstrate that they understand the need for a follow up, they also seem to be rather helpless. See, for example, the following recommendation: ‘The Council asks the Presidency to prepare a letter on the basis of the conclusions of the evaluation report on each Member State and to forward it to the Member States according to a timetable reflecting the original order of evaluations; each State would then have to describe the institutional, legal, practical, administrative and logistical measures it had taken or will take in response to the recommendations

²⁵⁴ One early example: The Council heard a presentation by the Commission on the biannual update of the scoreboard to review progress on the creation of an area of freedom, security and justice in the European Union. See Council of the European Union, Press Release 15691/02 (Presse 404), 2477th Council meeting Justice and Home Affairs, Brussels, 19 December 2002. That was contained in an 89 page Communication from the Commission.

²⁵⁵ We have been unable to trace the Member States’ practice on reporting back to the Council.

²⁵⁶ G. Vernimmen-Van Tiggelen and L. Surano, *Analysis of the future of mutual recognition in criminal matters in the European Union*, Université Libre de Bruxelles 2008, p.54.

²⁵⁷ The Commission has often referred to this possibility. See COM (2001) 771, p. 7.

²⁵⁸ The original deadline regarding this Framework Decision was 1 March 2003.

addressed to it. The outcome could then be passed on to the Council by means of a Presidency report.²⁵⁹

One may question whether Member States that do not have sufficient financial and human resources to implement legislation can be expected to have sufficient resources to live up to the findings of the evaluation mechanisms.

The Commission drew up reports on the implementation of the European Arrest Warrant.²⁶⁰ Additionally, the Council published a European handbook on it.²⁶¹ More generally, the Commission reports on the implementation of The Hague Programme by means of implementation scoreboards.²⁶² It does so by comparing the plans announced in The Hague Programme with the measures taken by the Union, as well as by looking at the implementation of legal instruments in the fields of justice, freedom and security at the national level.

The pressure put on Member States that are not complying takes the form of naming and shaming. However, one may question whether naming and shaming as an outcome is conducive to critical self-reflection, when reporting to the evaluation team. The Member State that failed to (fully) implement a measure is sometimes, but not always, specifically mentioned. Often, the Commission only expresses the view that the transposition of the legal instrument is generally unsatisfactory, or that Member States do not communicate information to the Commission at all.

With the exception of the Schengen mechanism, there is virtually no attention paid to a follow up of the evaluation outcomes. The structural nature of this is surprising, but at the same time it explains the lack of compliance. Why bother with compliance of outcomes if there is to be no follow up? Any evaluation of whatever scope and degree of institutionalization will seriously suffer from the absence of a follow up.

Only a few years ago, de Biolley and Weyembergh criticised the fact that the conclusions and recommendations were unanimously adopted by the Council, which raises questions about impartiality, but may be explained by the political and diplomatic sensitivity of the

²⁵⁹ Final report on the first evaluation exercise – mutual legal assistance in criminal matters, Brussels, 10 May 2001, 8648/01, Recommendation 21, p. 29.

²⁶⁰ See Report from the Commission on the implementation since 2005 of the Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between the Member States.

²⁶¹ Council of the European Union, Brussels, 8 November 2007, European handbook on how to fill in the European arrest warrant form. See also Draft Council resolution concerning an updated handbook with recommendations for international police cooperation and measures to prevent and control violence and disturbances in connection with football matches with an international dimension, in which at least one Member State is involved, Brussels, 24 October 2006.

²⁶² Report on the implementation of The Hague Programme for 2006 – Institutional Scoreboard – Table 1 and 2006 Implementation Scoreboard – Table 2, Brussels, 4 July 2007 11516/07 ADD 1 and ADD 2. See, in a more limited field: Follow-up to the Action Plan on trafficking in human beings, Brussels, 21 November 2006, 15321/1/06 REV 1; State of play of implementation by member States and EU bodies of EU priorities for the fight against organised crime based on the OCTA 2007, Brussels, 23 May 2008, 8102/3/08 REV 3.

evaluation.²⁶³ However, more recent initiatives demonstrate a different picture and urge a follow-up mechanism.²⁶⁴ Similarly, the Presidency has recently recommended follow up.²⁶⁵ Member States are asked to notify the actions they have taken since the evaluation.

Outside Schengen, a rather limited number of exceptions exist. In the field of combating terrorism, a follow up report was made. Member States report on the implementation of recommendations by the Council on counter-terrorism measures in the Member States.²⁶⁶ It was reported that 95% of the recommendations had been implemented. However, these estimates seem to be based on self-reporting, not on the independent collection of data. The Schengen evaluation system provides for follow up visits.²⁶⁷ The Schengen Executive Committee may make decisions, but may also give specific recommendations to States parties to the Convention Implementing the Schengen Agreement.

The follow up of evaluation reports regarding candidate Member States is more common, because they play a role in the final decision as to whether a State has complied with all requirements. In areas other than enlargement, not only are sanctions not provided, but the perspective is also not forward looking. De Biolley and Weyembergh phrase it very well: 'L'évaluation devrait donc être envisagée non pas comme un exercice ponctuel mais bien comme un processus évolutif, dynamique, à réaliser par étapes successives, a fin de renouveler les vérifications à intervalles réguliers.'²⁶⁸

National follow up

In the Netherlands, the government sent the peer evaluation report on the EAW to parliament.²⁶⁹ The government also added its comments to the report and the follow up it would undertake. The government clearly stated that only some of the recommendations would be followed, the reason being that the Framework Decision ought to be interpreted differently from the Commission, which is what the evaluation team did. The report and the government's response were discussed in parliament,²⁷⁰ and are published in the

²⁶³ Serge de Biolley and Anne Weyembergh, *L'évaluation dans le cadre du troisième pilier du traité sur l'Union européenne*, in: Weyembergh, A and S. de Biolley, *Comment évaluer le droit pénal européen?*, Bruxelles 2006, p. 86.

²⁶⁴ See, for example, the Benelux countries in *Gemeenschappelijke doelstellingen voor een nieuw EU-werkprogramma op het gebied van justitie*, in het kader van de ruimte van vrijheid, veiligheid en rechtvaardigheid. 09/AVT-NL/JU93450.

²⁶⁵ Final report on the fourth round of mutual evaluations – The practical application of the European Arrest Warrant and corresponding surrender procedures between Member States, Brussels, 18 May 2009, 8302/2/09, REV2.

²⁶⁶ See Follow-up report on the implementation of recommendations by the Council on counter-terrorism measures in the Member States. Brussels, 23 March 2007, 5356/2/07 REV 2.

²⁶⁷ R. Genson and W. van de Rijt, *L'évaluation de Schengen dans le cadre de l'élargissement*, in: Weyembergh, A and S. de Biolley, *Comment évaluer le droit pénal européen?*, Bruxelles 2006, p.217.

²⁶⁸ Serge de Biolley and Anne Weyembergh, *L'évaluation dans le cadre du troisième pilier du traité sur l'Union européenne*, in: Weyembergh, A and S. de Biolley, *Comment évaluer le droit pénal européen?*, Bruxelles 2006, p. 80.

²⁶⁹ There is no general obligation to do so under Dutch law. However, the Dutch implementation act provides for an evaluation report to be sent to parliament.

²⁷⁰ See 23490, nr. 545.

parliamentary documents. The evaluation report (hard copy) has been deposited in the parliamentary library.

We have been unable to extend our research into national follow up practices in other Member States. However, it can be noted that the Dutch practice on the EAW report complies with Article 70 TFEU, which also stipulates that national parliaments shall be informed of the content and results of the evaluation. As far as we can see, there is very little internal evaluation and follow up at the Member State level, with the notable exception being the United Kingdom. The House of Lords conducts evaluation hearings on several topics, both before EU legal instruments are adopted and afterwards.²⁷¹

Degree of institutionalization

Despite the fact that Joint Action 97/827 of 5 December 1997, which establishes a mechanism for evaluating the application and implementation at the national level of international undertakings in the fight against organized crime, has been the basis for numerous evaluations, it has not led to the establishment of a permanent infrastructure for evaluation. It is recommended that evaluation and monitoring be institutionalized.²⁷² *Ad hoc* evaluation teams must ‘reinvent the wheel’ and will lose expertise and effectiveness. With regard to various related topics, networks have already been established.²⁷³ Their goals seem to be to improve cooperation by better knowing each other.

The Schengen mechanism can be regarded as a more structural and institutional mechanism.

Effects and impact

With the exception of legislation in the field of terrorism, there is very little evidence that the evaluation mechanisms of the third pillar have been effective. The evaluation mechanisms applied can, by and large, be characterised as a bureaucratic exercise, for which the purposes are not (or, at best, no longer) clear. Evaluation has become a goal in itself, but it should not be.

The conclusion that the existing evaluation mechanisms are fragmentary and incoherent is, however, not new. The Commission itself reached this view already in 2006.²⁷⁴ The dissatisfaction of the Commission is linked to the fact that evaluations are limited in their scope to the implementation of a specific legal instrument. If one were to create a ‘topic based’ system of evaluation, more coherence could be developed, and factors not just related to implementation could be taken into account. This does include the evaluation of

²⁷¹ House of Lords, European Union Committee, 30th Report of Session 2005-06, European Arrest Warrant – Recent Developments, Report with Evidence, HL Paper 156.

²⁷² G. Vernimmen-Van Tiggelen and L. Surano, Analysis of the future of mutual recognition in criminal matters in the European Union, Université Libre de Bruxelles 2008, p.54.

²⁷³ Contactpunten ter bestrijding corruptie. European Crime Prevention Network.

²⁷⁴ Mededeling van de Commissie aan de Raad en het Europees Parlement, uitvoering van het Haags programma: koersbepaling, Brussel, 28.6.2006, COM (2006) 331definitief.

the legal instrument itself. In the absence of general studies on the follow up of evaluation findings,²⁷⁵ the impact of the existing mechanisms cannot be measured.

De Schutter characterized the peer evaluation mechanism set up by the EU as a combination of evaluation and collective learning.²⁷⁶ We interpret his concern in the sense that peer evaluations qualify quite well as a mechanism for mutual learning, but do not qualify well as a tool to be evaluated.²⁷⁷ It is important to maintain the mutual learning element, because otherwise law would be too static and it would be impossible to experiment with new methods.

5.2 Methodology of evaluation: more room for social scientists

Over the course of this research project, we came across vast mounts of literature on evaluation mechanisms outside the legal field,²⁷⁸ written by social scientists, predominantly criminologists. Although it would go beyond our expertise to give a complete overview of evaluation methodology, it is clear that there is a lot to learn from academics other than lawyers. Looking back at the efforts on evaluation and monitoring in the third pillar, the impression remains that lawyers without any expertise in evaluation have, indeed, been trying to ‘reinvent the wheel’.²⁷⁹ The quality of the evaluation, as well as the introduction of a methodology, could be seriously improved if this expertise would be taken into account.²⁸⁰ One of the most important lessons on evaluation to draw from the non-legal field is that data must be collected by an independent entity that does not have an interest in the outcome of the evaluation.

In its Communication on Evaluation of EU Policies on Freedom, Security and Justice,²⁸¹ the Commission proposes various steps. If one looks at the evaluation questions relating to Framework Decision 2003/577 on freezing orders, it is intended to count the number of orders issued and executed. One must seriously question whether such a method would result in any relevant data. The expected outcome is an increased efficiency and

²⁷⁵ Since there is no information on the follow up in Member States at the EU level, the information is scattered all over the Member States. It would be a Herculean task to collect that data in all Member States.

²⁷⁶ Olivier De Schutter, *The Role of Evaluation in Experimentalist Governance: Learning by Monitoring in the Establishment of the Area of Freedom, Security and Justice*, Faculté de Droit de L’Université de Louvain, November 2008, p. 3.

²⁷⁷ Olivier De Schutter, *The Role of Evaluation in Experimentalist Governance: Learning by Monitoring in the Establishment of the Area of Freedom, Security and Justice*, Faculté de Droit de L’Université de Louvain, November 2008, p. 19.

²⁷⁸ We limit ourselves to the standard work in evaluation of Peter H. Rossi, Howard E. Freeman and Mark W. Lipsey, *Evaluation, A Systematic Approach*, Sage Publications, Thousands Oaks – London – New Delhi 2006.

²⁷⁹ Criminologists are also desperately needed at the legislative stage. See André Klip, *Criminal Law in the European Union*, Inaugural Lecture Maastricht University 2004, p. 54 and C. Fijnaut, *De Europese Unie: een luthof voor strafrechtelijke rechtsvergelijking*, Deventer 2001.

²⁸⁰ Nelen mentions, for example, that researchers on evaluation are often instructed by their evaluation sponsors not to include the legislation itself within the scope of the evaluation. However, evaluation should be broader. See J.M. Nelen, *Gelet op de wet. De evaluatie van strafwetgeving onder de loep*, Diss. Free University of Amsterdam 2000, p. 176-179.

²⁸¹ Communication from the Commission to the Council and the European Parliament, *Evaluation of EU Policies on Freedom, Security and Justice*, COM (2006) 332 final, Brussels, 28.6.2006.

effectiveness of freezing orders, measured by, among other things, quicker procedures to freeze assets. That would require making a comparison with the situation before the Framework Decision entered into force.

To measure the impact raises even more methodological questions. It is expected that using more freezing orders will decrease the levels of organised crime. Apart from the fact that it might be extremely difficult to measure the level of organised crime at all, it is highly unlikely that any increase in the number of freezing orders will itself be the significant factor in reducing it. In relation to other instruments, there is the clear presumption that transposition of obligations will have an effect on the crimes committed by citizens. From the criminological literature we have seen, it is generally known that criminal legislation as such only has very little influence on the reduction of crime.

Various evaluation policy questions are raised in the literature,²⁸² which are relevant, but not always referred to in the context of EU evaluations:

- How often should this programme be evaluated?
- Who determines the focus of this evaluation?
- Who is responsible for doing this evaluation?
- How much should we budget for this evaluation?
- What evaluation resources and support should the system supply?
- How much time should people spend on this evaluation?
- Who is responsible for constructing and managing the evaluation databases?
- Should we utilise an internal or external evaluation for this program?
- Should we use a controlled or comparative design to evaluate this program?
- How should the evaluation results for this programme be disseminated?²⁸³

Some of these questions appear in the evaluation, other do not. In relation to the follow up, evaluation literature also offers some lessons.²⁸⁴ It is important to specify the follow up, and what is to be expected from it, already at the moment of establishing the whole evaluation.

5.3 Lessons learned

Repeating the conclusion that current evaluation mechanisms are fragmentary, inconsistent and implementation based, we subscribe to the conclusion of the Commission that there is an urgent need for improvement of the evaluation mechanism.²⁸⁵ The lessons learned from our analysis focus on the following categories:

²⁸² See, for example, Nelen, p. 9-22.

²⁸³ Based on William Trochim, Introduction to Evaluation, <http://www.socialresearchmethods.net/kb/intreval.htm>, last accessed 28 July 2009.

²⁸⁴ Nelen, p. 143-172.

²⁸⁵ See COM (1999) 752.

the quality of the legislation; the scope of the evaluation mechanism; its method; the way a follow up is ensured; and the degree of institutionalization of the mechanism.

We consider each of the elements of quality, scope, method, follow up and institutionalization as the most relevant in order to determine the contents and effectiveness of the system and thus factors necessary for its improvement. We are aware that these five categories are also intertwined, in the sense that, for example, if the scope of the evaluation mechanism is limited to evaluating the implementation in the Member States of a specific legal instrument, a permanent institution is not absolutely necessary. After having dealt with each individual category, we will present some recommendations for each relevant element.

The quality of legislation

A link between the results of evaluation and the general results on compliance with binding norms has previously been identified.²⁸⁶ Whilst many recommendations resulting from evaluation relate to the unique make-up of specific Member States, some common issues may emerge during the findings and recommendations. These should be used.²⁸⁷ If many Member States can honestly report that there is no necessity to implement legislation, perhaps because their existing legislation is already in compliance with the new EU instrument, this raises the question as to why the instrument was necessary in the first place.²⁸⁸ Vague norms, which may leave considerable discretion to the Member States, lead to disputes about compliance with both the implementation obligation, as well as with the follow up to evaluation findings. This can partly be solved by providing for infringement procedures and, to a large extent, can be prevented by clearer texts.²⁸⁹

It is recommended that:

- **in the drafting process of EU legislation, Member States report on whether implementation would be necessary for their national criminal justice system. This may mean that EU legislation is only adopted when necessary.**

²⁸⁶ See, for example, Final report on the fourth round of mutual evaluations – The practical application of the European Arrest Warrant and corresponding surrender procedures between Member States, Brussels, 18 May 2009, 8302/2/09, REV2, p.7: ‘The findings of the evaluation demonstrate, however, that in some Member States (...). As a matter of principle, this situation seems difficult to reconcile with the letter and the spirit of the Framework Decision, irrespective of how understandable it may be in view of the specificities of the national system or associated practical advantages.’ Likewise, some of the recommendations of the report directly relate to compliance with implementation. See, for example, Recommendation 17.

²⁸⁷ Final report on the fourth round of mutual evaluations – The practical application of the European Arrest Warrant and corresponding surrender procedures between Member States, Brussels, 18 May 2009, 8302/2/09, REV2.

²⁸⁸ Weyembergh and Santamaria, Conclusions, in: A. Weyembergh and V. Santamaria, The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings, Bruxelles 2009, p. 391.

²⁸⁹ Stine Andersen, Sovereignty and the emergence of non-binding peer review within the EU, p. 11.

- **Vague standard setting should be avoided, since, although it may have the appearance of harmonisation, it has adverse effects in practice.**

Scope of the evaluation

The evaluation mechanism is used in a fragmentary and incoherent manner. Some legal instruments are looked at individually. The focus is predominantly on implementation.²⁹⁰ This perspective should be broadened. Evaluation exercises should not be automated. Each evaluation requires thorough preparation regarding what should be evaluated, how this should be done and by whom. Evaluations should aim at obtaining information on how to make a certain set of rules better.

Respect for the rule of law must be considered as a condition for compliance with evaluation outcomes. The choice of the topic to be evaluated must be made at the political level. However, the actual evaluation should preferably be performed by an independent entity. One can easily subscribe to the Council Conclusions of 19 June 2007: ‘The Council considers that the following criteria should apply when selecting the areas for evaluation: (1) availability of data for the area concerned, (2) avoidance of overlap with the executing evaluation mechanisms for individual legal acts, (3) no duplication of other Commission evaluation activities, e.g. in the sectors of justice policy and criminal matters, and (4) relevance of the area selected for freedom, security and justice policy goals.’²⁹¹

In its recent Communication on the Stockholm Program, the Commission stated that: ‘there has to be evaluation of the effectiveness of the legal and political instruments adopted at Community level. Evaluation is also necessary to determine any obstacles to the proper functioning of the European judicial area. Evaluation should be carried out periodically, and should facilitate better understanding of national systems in order to identify best practice.’²⁹²

Broadening the scope is, arguably, also necessary in order to inform the Commission about the general legal context in the Member States.²⁹³ It will thus contribute to further expanding the expertise of the Commission. The European Parliament has recommended that evaluation ‘should assess the impact of EU policies on the ground and on the daily management of justice, as well as the quality, efficiency, integrity and fairness of justice,

²⁹⁰ Whether a Member State may, in substance, reach the result by other means, has not been considered. See Weyembergh and Santamaria, Conclusions, in: A. Weyembergh and V. Santamaria, *The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings*, Bruxelles 2009, p. 388.

²⁹¹ Presidency, *Evaluation of EU policies on Freedom, Security and Justice - Council Conclusions*, Brussels, 19 June 2007, 10893/1/07.

²⁹² Brussels 10 June 2009, COM (2009) 262/4, *An area of freedom, security and justice serving the citizen*, p. 11.

²⁹³ Serge de Biolley and Anne Weyembergh, *L’évaluation dans le cadre du troisième pilier du traité sur l’Union européenne*, in: Weyembergh, A and S. de Biolley, *Comment évaluer le droit pénal européen?*, Bruxelles 2006, p. 85.

also taking into account the level of implementation of ECJ and ECHR case-law by Member States.²⁹⁴

The following recommendations are made regarding the scope of the evaluation:

1. The **evaluation should not be limited to implementation only**, but also take into consideration the instrument itself, as well as the consequences for the criminal justice systems of the Member States (which includes both the quality of the instrument, as well as its effectiveness).
2. **Evaluation should be forward looking** and include an analysis of how to make the instrument ‘better’.
3. There should be **less focus on what is wrong with a specific Member State**, and more on what could be improved.
4. **Evaluation should be topic based**, instead of legal instrument based - for example, a Member State’s respect for the rule of law. This enables a more coherent picture to be ascertained. Or, in the words of the Commission: ‘priority setting based on risk analyses’²⁹⁵.
5. **Internal Member State factors determining (non-) compliance should be looked at**. They are relevant for a contextual approach.

The methodology

The methods used for the various mechanisms in peer evaluation are relatively consistent. Questionnaires are drawn up and completed by the Member States. Peer evaluation teams are set up to perform a so-called mutual evaluation. The outcome is not always comparable, nor is it always shared among the Member States or even published. The quality of the report depends on the quality of the evaluation team and the willingness of the Member States in question.²⁹⁶

Evaluation/monitoring often takes place at a time when there is hardly any experience in practice with the legal instrument. This reinforces its focus on implementation, because there is not much else to look at. It is doubtful whether the evaluation teams have access to a suitable benchmark for the purposes of comparison. Expertise and knowledge is not retained within the team and not transferred onto the next team. Knowledge from the

²⁹⁴ EP Recommendation of 7 May 2009 to the Council on development of an EU criminal justice area (2009/ 2012 (INI)). See also Jegouzo, p. 153.

²⁹⁵ Decision on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 105 final, 2009/0032 (CNS) and Regulation on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 102 final, 2009/0032=3 (CNS).

²⁹⁶ Serge de Biolley and Anne Weyembergh, L’évaluation dans le cadre du troisième pilier du traité sur l’Union européenne, in: Weyembergh, A and S. de Biolley, Comment évaluer le droit pénal européen?, Bruxelles 2006, p. 96.

social sciences is not used. As a whole, the current third pillar evaluation mechanism can be characterized as a missed opportunity.

The methodology used by the Commission for its reports on implementation simply consists of a rather straightforward legal comparison of the national implementing legislation and the relevant EU legal instrument.

There is no visible attempt to harmonise the various mechanisms of evaluation. However, in the context of the Schengen mechanism, the Commission proposed the following in its 2009 drafts for a new Regulation and Decision:

- ‘1. The current methodology for the evaluation mechanism is inadequate. The rules on consistency and frequency of evaluations are not clear. There is no practice of conducting unannounced on-site visits.
2. There is a need to develop a methodology for priority-setting based on risk analysis.
3. A consistently high quality of expertise during the evaluation exercise needs to be ensured. The experts participating in the evaluation should possess an adequate level of legal knowledge and practical experience. Sending an expert from each Member State on each on-site visit may be detrimental to the efficiency of the exercise. An appropriate number of experts to participate in visits needs to be determined.
4. The post-evaluation mechanism for assessing the follow-up given to recommendations made after the on-site visits needs improving, as the measures taken to remedy deficiencies as well as the timeframe within which they are to be remedied vary from one Member State to another.
5. The institutional responsibility of the Commission as guardian of the Treaty concerning first pillar matters is not reflected in the current evaluation system.

The following points are intended to address these weaknesses:

Methodology for evaluations

The present proposal introduces a clear programming, providing for multiannual and annual programmes of announced on-site visits. Member States will continue to be evaluated on a regular basis in order to ensure the overall correct application of the *acquis*. All parts of the Schengen *acquis* which have their legal basis in the Treaty on European Union can be the subject of evaluation.

This evaluation can be based on replies to questionnaires, on-site visits or a combination of both. In the latter case, the visits can take place shortly after the replies to the questionnaires are received.

In recent years, Member States have not seen the need to carry out evaluations on the spot concerning judicial cooperation in criminal matters and drugs. Data protection has also not always been subject to on-site evaluations.

Nevertheless, on-site visits are not limited to police cooperation and can cover all parts of the Schengen acquis. The concrete need for such visits will be determined by the Commission after seeking the advice of the Member States, taking into account changes in the legislation, procedures or organisation of the Member State concerned.

In addition, if necessary, thematic or regional evaluations can also be included in the annual programme.

Both multiannual and annual programmes can always be adapted if need be.

As regards the methodology for priority-setting based on risk assessment in relation to the Schengen acquis on police cooperation and on judicial cooperation, the Commission does not yet see a role for Europol and Eurojust given their current mandates. However, if Europol or Eurojust produce risk analyses for the acquis in question, the Commission will certainly use these in order to plan the evaluations.²⁹⁷

Both the Commission and the Dutch Ministry of Justice identified three stages:

- a. independent findings;
- b. analysis and recommendation for action;
- c. supervision of the Member States' reaction and compliance.²⁹⁸

However, before the evaluation is to be carried out, a more elaborated plan must be developed. From a brief look at the literature on evaluation, we conclude that it is important to discern various stages. Nelen presented a model from a criminological perspective.²⁹⁹ Rossi and others argue that 'every evaluation must be tailored to its program.'³⁰⁰ It is absolutely necessary to include criminologists and other experts in any assessment of the establishment of a new evaluation mechanism.³⁰¹ A model for an academic evaluation has been prepared by the European Criminal Law Academic Network (ECLAN).³⁰² Information from sources other than the Member States should be admitted in a structured way.

²⁹⁷ Decision on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 105 final, 2009/0032 (CNS) and Regulation on the establishment of an evaluation mechanism to monitor the application of the Schengen acquis, Brussels, 4.3.2009, COM (2009) 102 final, 2009/0032=3 (CNS).

²⁹⁸ Contribution for a Mutual Trust Chapter of the Stockholm Programme, Note prepared by the Directorate of International and European Affairs of the Ministry of Justice. The Hague, 28 May 2009. Proposals regarding the focus and set-up of an additional evaluation programme in the field of EU judicial cooperation in criminal matters. This three step approach resembles the approach of the Commission in Communication from the Commission to the Council and the European Parliament, Evaluation of EU Policies on Freedom, Security and Justice, COM (2006) 332 final, Brussels, 28.6.2006, p. 7.

²⁹⁹ Nelen, p. 186-189.

³⁰⁰ Peter H. Rossi, Howard E. Freeman and Mark W. Lipsey, *Evaluation, A Systematic Approach*, Sage Publications, Thousands Oaks – London – New Delhi 2006, p. 37.

³⁰¹ See, for example, the report being prepared at request of the Netherlands Ministry of Justice by Maastricht University on the evaluation of Dutch anti-terrorism measures. F. Leeuw, H. Nelen, E. Wolff and S. Luthuli, *Antiterrorismebeleid en evaluatieonderzoek: Framework, toepassingen en voorbeelden* (forthcoming).

³⁰² A. Weyembergh and V. Santamaria, *The Evaluation of European criminal law. The example of the Framework Decision on combating trafficking in human beings*, Bruxelles 2009.

The following recommendations on the method can be made:

1. **Evaluations should make use of the expertise of evaluations studies**, in order to draft an appropriate methodology;
2. **Evaluations should take place later** (after full implementation and some years of practical use);
3. **The possibility of unannounced on-site visits should be considered**,³⁰³
4. **Information should be obtained from sources other than the Member State**;
5. **Independent members**, experts not being peers, should be on the evaluation team,³⁰⁴
6. **Empirical data** should be used, not just a comparison of legal texts;
7. A **permanent structure** should be established (instead of ‘reinventing the wheel’),³⁰⁵
8. The team should include non-lawyers, particularly criminologists;³⁰⁶
9. Findings and outcomes of evaluation reports are non-negotiable.

Follow up

The response to evaluation mechanisms should at least be obligatory. This might be stimulated by findings of non-compliance before the Court. However, it seems wiser to attempt to end situations where either an unsatisfactory follow up, or even no report at all, is the final result of particular legal instruments. A general legal instrument on evaluation should provide further for visits (possibly unannounced), in order to force Member States to follow up. It should be possible to pay more attention to one Member State rather than another Member State, depending on the status of compliance. Similar

³⁰³ See for Wolfgang Rau and John Ringguth, Evaluation in the framework of the Council of Europe, in: Comment évaluer, p. 33-44.

³⁰⁴ Also advocated by H. Nilsson, Eight years of experiences of mutual evaluation within the EU, in: Comment évaluer, p. 124 and O. de Schutter and Valérie van Goethem, The added value of a systematic and regular monitoring of the situation of fundamental rights in the Member States for the evaluation of the implementation of Union laws and policies, in: Comment évaluer, p. 125-145.

³⁰⁵ Of course, this is linked with the issue of institutionalization.

³⁰⁶ We require statistics on which to base conclusions and findings, when making decisions on legislation. See also Communication from the Commission to the Council and the European Parliament, Evaluation of EU Policies on Freedom, Security and Justice, COM (2006) 332 final, Brussels, 28.6.2006, p. 6.

to other areas, it might be possible to provide technical advice to Member States that have difficulties in complying with the findings of evaluation mechanisms.³⁰⁷

Therefore, the following recommendations are made:

1. **A general legal instrument on evaluations should be drafted, containing the obligation that evaluations should be followed up;**
2. The **status of the evaluation should be clear** from the start: should recommendations be implemented? If yes, how are they followed up? If no, what then is the purpose of the exercise?
3. Visits should take place at a reasonable time after the findings of the evaluation are presented. They should be repeated if necessary.

Institutionalization

There is little institutionalization in the sense that there is no permanent committee performing evaluations and building up and maintaining expertise, although there are some consistencies in either the Commission or the Council almost always being a member of evaluation teams. However, that does not make up for the fact that every evaluation starts from scratch and that individual members are often different. This is linked to the limited scope of the evaluation: one legal instrument for each legal evaluation.

This does not promote comparable standards of evaluation and does not create an overview of the general interests. There are calls for more structural approaches towards monitoring.³⁰⁸ An evaluation network has been proposed by the European Parliament.³⁰⁹ Monitoring should be performed by (treaty based) agencies with established membership, responding to strict eligibility criteria and by independent bodies.

It is recommended that:

1. **A permanent structure for evaluation and monitoring** be established;
2. **The same team should evaluate all Member States.**

³⁰⁷ De Schutter mentions that for the Fundamental Rights Agency. See O. De Schutter, *The Two Europes of Human Rights: The Emerging Division of Tasks between the Council of Europe and the European Union in Promoting Human Rights in Europe*, 14 *The Columbia Journal of European Law* 2008, p. 523.

³⁰⁸ See, for example, in relation to the setting up of a methodology Report on the practical operation of the methodology for a systematic and rigorous monitoring of compliance with the charter of fundamental rights, Brussels, 29.4.2009, COM(2009) 205 final.

³⁰⁹ EP Recommendation of 7 May 2009 to the Council on development of an EU criminal justice area (2009/ 2012 (INI)).

6. Synthesis and recommendations

In this report, we have analyzed, based upon existing practices, those factors that hinder (better) compliance (and follow up) with monitoring and evaluation mechanisms in the first and third pillars of the EU. We based our findings on an in-depth analysis of existing literature on this topic (see the Annex to this report for a schematic overview, as well as chapters 2 and 3 for descriptions), as well as on evidence based analysis of real-life cases and experiences with monitoring and evaluation mechanisms (see chapters 4 and 5).

This allowed us to learn from the experiences with monitoring and evaluation mechanisms in a wide variety of situations, such as in the OECD, the first pillar and the third pillar of the EU (see, in particular, 4.5 and 5.3). This concluding section will, in particular, concentrate on providing a more schematic representation of the various lessons to be learned. We start with the most ‘abstract’ lessons related to contextual factors, and conclude with more practical lessons related to the functioning of monitoring and evaluation mechanisms *per se* in various scenarios.

6.1 Lessons learned I – From the more abstract context...

The main – and very obvious – lesson is the realization that a well-functioning monitoring or evaluation mechanism needs a **context sensitive approach**, rather than a one size fits all or implementation biased approach. A context sensitive approach needs to be cognizant of the following two crucial elements:

1. Monitoring and evaluation mechanisms, to a large extent, stand or fall with the degree to which its members learn from each other and thereby adhere to the rules, which in turn depends upon the **level of commitment** of its members. Learning from the examples provided by, amongst others, the OECD (see 3.1) and the EES (see 4.1-4.4), as well as from anti-terrorism measures, we see that learning and/or compliance is/are highly dependent upon the willingness of the members to participate. This level of commitment or willingness is most likely to be found in a situation of shared views and values, trust and perceived legitimacy. In other words, in order to stimulate compliance or follow up with monitoring and evaluation mechanisms in the third pillar, the EU needs to ensure that its Member States perceive the legislation and the monitoring and evaluation mechanisms as legitimate, fair and necessary. As we learn from the rationalist perspective (see 2.4 and the Annex to this report), at first sight countries all act as rational actors who conduct cost-benefit calculations; in order for participating countries to comply with a monitoring or evaluation mechanism, they either need to be convinced of the ‘benefits’ of compliance, or they need to have faith or trust in the political environment (in other words, they need to know that when they comply with the requirements, countries x, y and z will also do so in order to avoid unequal circumstances). Only in a ‘trust-relation’ will the participants in a monitoring or evaluation mechanism potentially feel committed to the process at hand. This could, for example, be stimulated if the topic of evaluation is selected more carefully and not simply as an automatic consequence of legislation.

Therefore, for every monitoring or evaluation mechanism that is to be established at the EU level, we carefully need to **examine the political will amongst Member States to achieve the required goal**. Without a basic level of trust and commitment among the members, even the best designed evaluation or monitoring mechanism is subject to failure.

2. A second necessary element in order to allow for a context sensitive approach is the realization that specific Member State factors do matter and that these domestic factors are, generally speaking, rather ‘sticky’ or ‘culturally determined mental patterns’, and thus difficult to overcome. The overview of factors that explain compliance problems (see also the Annex to this report) demonstrates how crucial it is to realize that there is not one single factor that explains the lack of compliance. In different situations, and for different Member States, different factors are at play. In some circumstances, or for some Member States, it will be more effective to generate learning incentives, whilst for others it is more beneficial to threaten the imposition of sanctions. This suggests that several evaluation methods may not necessarily be mutually exclusive. Peer review exercises involving Member State representatives and national independent experts may well be operating alongside ‘hard’ sanctions. Unfortunately, it is impossible to divide all situations or all Member States into clear cut categories. Domestic circumstances change over time – governments change, economic circumstances change, the relevance and importance of certain societal actors or policy sectors change – and thus it is impossible to identify which countries, and in what situations require an EU level response from a structural point of view. Evaluation and monitoring mechanisms therefore need flexibility to allow for the relevant actors to become aware of, and act on these crucial domestic specificities that so much influence how Member States respond to EU level requests to improve compliance or follow up. Additionally, because not all Member States find themselves in equal circumstances, the follow up may differ from Member State to Member State, according to their needs, and may also amount to offering specific assistance. **Actors that play a dominant role in ensuring follow up in these mechanisms need to be(come) sensitive to these domestic specificities and need to have the freedom, and be able, to react upon them.**
3. **A third element, which particularly comes to the fore in the field of criminal law, is that the interdependence of Member States and exclusive nature of the legislation, stimulate compliance at all levels**, both in relation to compliance with legislation, as well as with the outcomes of evaluation findings. The EAW and the anti-terrorism mechanisms are good examples of this. Both instruments can only work if all Member States perform well- they are interdependent. With regard to the EAW, this is further stimulated by the exclusive nature of the mechanism. A Member State not implementing and/or not complying with the evaluation reports may not be able to use the mechanism. Terrorism is a common threat to all Member States. Member States must therefore be aware that they depend on each other. A spirit supportive of not only the instrument, but also its EU nature and cooperation with other Member States, will stimulate compliance.

This could also be promoted by a common policy on certain areas, which is not only formally or symbolically supported by the Member States, but also appears from what they actually do.

6.2 Lessons learned II – ...to the more pragmatic functioning of evaluation and monitoring mechanisms – possible scenarios

When thinking of monitoring and evaluation mechanisms in terms of varying degrees of ‘strictness’, we can identify a scale of softer and harder mechanisms currently in place. While the potential number of ‘steps’ or ‘scenarios’ in such a scale is much longer, we would – for the sake of simplicity – like to focus on three scenarios, based on different logic, which may, nonetheless, co-exist in practice. What conditions are at play, gleaned from the case studies described above, which influence the success or failure of each of these scenarios, and should therefore be taken into consideration when designing effective mechanisms that aim to enhance compliance?

1. Monitoring and evaluation mechanisms based on peer review

On a positive note, those monitoring and evaluation mechanisms that largely rest on peer review among the member countries – think of the examples within the OECD or the EES – may be beneficial in stimulating mutual learning via peer pressure. Member States seem to be relatively open to accepting peer review from other Member States, and the intensive, non-confrontational and non-legalistic nature of the interaction in most peer review systems generally leads to an ongoing and informal information sharing experience, from which all partners may benefit. One of the major downsides with regard to traditional peer review mechanisms is their self-reporting aspect. Countries may have their motivations to refrain from reporting truthfully, with all its negative consequences. In addition, given the lack of ‘hard sanctions’, there may be only weak incentives for Member States to implement policy guidelines and/or policy recommendations.

2. Monitoring and evaluation mechanisms based on independent, ‘expertocratic’ reporting

In response to the downsides of self-reporting mechanisms, a logical step would be to establish monitoring and evaluation mechanisms based on independent reporting. It is argued that, only if reporting is done in an independent, uniform, objective and standardized way, is it possible to arrive at a genuine comparability of national policies and of their respective impacts – something that is necessary for effective learning, and for the Commission to effectively perform its monitoring and evaluation functions. While independent experts might be better placed to report accurately and truthfully on the Member States’ performances, the case study on the anti-discrimination field (see 4.6) demonstrates that truly independent reporting can only take place when there are adequate resources, and a sufficient degree of independence from government, for the reporting bodies to actually function as expected. However, even if the virtues of independent reporting are numerous and obvious, it is not clear that Member States will readily accept independent experts to step in and expose what might be sensitive

information. Of course, this readiness to accept expert reporting is also very much a matter of commitment to making the evaluation mechanism work (see 6.1).

3. Monitoring and evaluation mechanisms based on sanctioning

Moving one step further, we can consider the use of sanctions to stimulate (by imposing fines) or to force (through the launch of infringement proceedings) Member States to comply. However, while the presence of clear sanctioning mechanisms is a positive incentive for Member States to comply – particularly in research comparing the EU to other international organizations, given the relatively strong role of the Commission and ECJ – we should not overestimate its influence. Monetary fines may not always work effectively and, based on the information provided in 2.3, we also note that the EU's infringement procedure does not always work smoothly. It is, for example, a better instrument for undoing transposition problems than for tackling practical implementation difficulties. Moreover, a greater resort to strict sanctioning in the third pillar requires the necessary political commitment to make the mechanism work in practice. The failure to sanction France and Germany in the field of the Growth and Stability Pact (see 4.7) is a telling example. If this lack of a firm political commitment on the part of the Member States is highly problematic, the discretion granted to the Commission in whether or not to start or continue an infringement procedure may also give rise to problems of enforcement. For example, decisions to move from the formal notice stage to a reasoned opinion must be taken by simple majority of the college of Commissioners. This implies that such a decision making process can easily become a political game, with national interests at stake, rather than a rational and objectively founded procedure according to which 'punishing' takes place fairly and effectively.

6.3 To conclude: achieving uniformity while respecting diversity

No matter which logic, or combination of logic, a monitoring or evaluation mechanism follows³¹⁰ – peer review, independent reporting and/or sanctioning – similar problems appear to arise everywhere. In all cases, we are met with difficulties relating to a lack of data, variability of data and problems of comparability of data. While at first sight it appeared that these problems were largely related to the fact that, in traditional peer review mechanisms, Member States themselves supply the data, the case studies in the anti-discrimination field, as well as the analysis of the European Commission as the guardian of the Treaties, demonstrate that all types of actors – Member States, independent experts, the European Commission – experience difficulties related to the collection of reliable and comparable data. However, it is self-evident that, in order to enable evaluation mechanisms to function effectively, there must be a streamlining and standardization of data collection methods across the board. Ideally, the process of data collection should be undertaken according to the following conditions.

Data collection is most likely to result in reliable and comparative information, which is crucial for checking compliance and stimulating follow up, when:

³¹⁰ We should note here that, in reality, we hardly ever experience such a strict division, and that different logical bases are usually involved.

- the monitoring or evaluation process is **institutionalized** (Member States are more likely to participate in, and accept the outcomes of, institutionalized mechanisms, since institutionalisation creates a dynamic of its own);
- the monitoring or evaluation process takes place on the basis of an **iterative process** supported by a **permanent structure** (if Member States are aware that their performance is judged regularly and repetitively, the probability that they will feel compelled to continue paying attention to the issue(s) discussed is greater than where *ad hoc* evaluations are concerned, due to the follow up element that it implies);
- the monitoring or evaluation process uses **common European, precise indicators for assessment** (it is self-evident that progress can only be measured comparatively if there are precisely defined and explicit indicators that limit the leeway for interpretation for those who report on national progress);
- the monitoring or evaluation process results in **prescriptive and express policy recommendations** that allow for effective monitoring and evaluation, and that give the process a forward-looking orientation. Only with such recommendations can the evaluation process be constructive and enhance compliance, while preventing the emergence of evaluation fatigue.

Without necessarily indicating a preference for a particular scenario or a specific combination of scenarios, we would emphasize that any system that intends to promote compliance with the implementation of EU agreements, as well as compliance with monitoring and evaluation results, needs to adhere to a **multi-step approach**. Without precisely specifying the number of steps needed, an analysis of compliance outcomes in the first and the third pillars of the EU seems to at least indicate the need for an inventory step, where information is collected in the different Member States in the most uniform and objective manner (see recommendations above). In order to ensure actual compliance with monitoring and evaluation outcomes, we would, however, support the shift towards a more flexible approach, reaching beyond an initial uniform and generally applied methodology; an idea based on our general conclusion that well-functioning monitoring and evaluation mechanisms need a context sensitive approach that takes into consideration political will and commitment, as well as domestic circumstances.³¹¹ Indeed, ensuring compliance and follow up cannot be achieved without creating some flexibility to react to these different – and changing – domestic circumstances, together with the accompanying different responses to ‘EU-interference’. This, combined with the fact that non-compliance is never caused by one single factor that is equally applicable to all Member States, demonstrates that we cannot assume that a single response to non-compliance from the EU level will have the same optimal results in all Member States. We thus suggest a system that acknowledges these circumstances, and allows for variation in reacting to non-compliance in different cases and in different Member States: sometimes we need to resort to the ‘stick’, but at other times, it is better to use a ‘carrot’.

³¹¹ As an example, refer to the follow up on respect for the rule of law in Romania and Bulgaria.

Annex – Overview of compliance factors

Dominant factors influencing compliance with EU legislation in the First and Third pillars

1. Factors related to the nature of the legislation / international regime

Factor	Hypothesis /Explanation	Lessons Learned
Nature / quality of the legislation	Highly detailed, complex legislation leads to compliance problems Vague / compromised legislation leads to interpretation problems, and thus to compliance problems ³¹²	Pay attention to the quality of legislation in the drafting phase Provide more guidelines, for example, EU-wide minimum criteria for inspection tasks, minimum standards for handling complaints and guidelines on access to national courts
Fairness / legitimacy of the legislation	The less the legislation is perceived as fair in the eyes of the country, the less likely it will be that compliance will take place	Providing possibilities for affected parties to interact in the decision making process increases the perception that the rules are legitimate The use of impact assessments during policy formation will increase the likelihood of an acceptance of the legitimacy of the rules
Ownership of the legislation	The fewer Member States that understand the problem that the legislation seeks to address, the less likely it will be complied with	Conduct problem analyses with a social sciences methodology (particularly criminological research) on the relevance of/need for the legislation
Enforcement/sanctioning capacities of the international organization ³¹³	The stronger the official enforcement capacities within an international organization (for example, the strong role of the ECJ in the EU, in comparison to other international organizations) - that is, the greater sanctioning power that it has - the more likely it is that its members will comply	Pay strong attention to the monitoring and sanctioning capacities of the international organization

³¹² These two hypotheses are somewhat contradictory. Evidence is found in both cases: directives that are too detailed can lead to compliance problems, but directives that are too vague are also shown to have these problems.

³¹³ A major downside of this hypothesis is that it does not explain the variation in compliance levels within the EU, but it is more applicable to comparing the EU with other types of international regimes.

Transparency within the international organization	The more information is available on the compliance behaviour of all participating countries, the more likely an individual country will itself comply	Make use of naming and shaming / scoreboards, etc. – these provide transparency, which may stimulate compliance by Member States
--	--	--

2. Factors at the level of the EU (Monitoring/Evaluation) (mainly related to Third pillar criminal law)

Factor	Hypothesis/Explanation	Lessons Learned
Nature of norms	The willingness to comply with evaluation findings corresponds to a willingness to comply with binding norms	Pay attention to the formulation of binding norms: make sure that the norms are clear, so as to avoid misinterpretations and (possibly) eventual non-compliance (see also below)
Exclusiveness of the legislation	The more exclusive EU legislation is, the better MS will comply with both the legislation and the evaluation mechanisms (see, for example, the European Arrest Warrant)	Encourage EU legislation to completely replace the previous system in place
Infringement proceedings	A lack of infringement proceedings (a lack of possibilities for enforcement of compliance) leads to a lessening concern of the MS to comply with evaluation findings	Consider the option of introducing stronger enforcement mechanisms, for example, the possibility of the Commission initiating infringement proceedings in relation to all legislation
Standard setting	When the binding norm is vague, evaluation of compliance may lead to a dispute as to the interpretation of the norm The moral authority of the Commission on interpretation of standards is weak	Avoid vague binding norms: having vague norms raises the question of why we have this legislation at all, if no agreement can be reached on more concrete norms? Avoiding vague norms also limits the possibility of diverging interpretations Avoid the appearance of the Commission as usurping the position as the highest authority of interpretation
Frequency/length of monitoring processes	A multiplicity of monitoring operations results in evaluation/monitoring fatigue: evaluation of	Limit evaluation and monitoring to specific, concrete areas: this will produce more reliable

	implementation being performed on the automated pilot and MS not really caring about the evaluation results	findings and will be conducive to the formulation of more realistic recommendations to which attention will more likely be paid by the addressees
Sharing of the outcome of evaluations	If MS do not know the evaluation results and recommendations made to other MS, they will not have the chance to change their behaviour before they are themselves evaluated, which reduces the efficiency and effectiveness (for example, missed opportunities for policy learning) of the evaluation mechanism	Publicise the evaluation/monitoring results to all MS and to the relevant stakeholders
Implementation-bias	Focusing just on Member States' compliance with transposition obligations is too narrow an enterprise: by merely checking compliance with the 'law in the books', rather than examining transposition, application and enforcement, the goal of 'practical compliance' may be left out	Broaden the scope of evaluation exercises to also include the stages of application and enforcement of the transposed legislation
Sanctions	An absence of sanctions for non-complying with the outcome of monitoring/evaluation mechanisms does not encourage compliance	At the least, create pressure on laggards by publicising the evaluation findings to all MS Preferably, create public pressure on governments by also publicising results to the public
Evaluation teams	As differently composed teams evaluate the MS, diverging standards of comparison are applied in practice, something that may not yield optimal results in terms of comparability of results and perceived legitimacy of the evaluation mechanism	Establish one evaluation team for all MS
Grounds for refusal in Framework Decisions on mutual recognition	Having this grounds of refusal provision undermines mutual trust (since application of this principle requires an investigation into whether these grounds apply and whether the MS deserves to be trusted), and creates differences of opinion between MS and evaluation teams	Abolish the grounds of refusal provision as far as possible

Standards for criminal procedures	As instruments in criminal law are assessed individually and are not tested against the background of a general framework, the cohesion of the criminal justice system is compromised	Formulate a common view on the role of the EU in the area of criminal law, and on the role of criminal law in the EU
EU view on rights of the individual	The lack of a clear view on the rights of the individual <i>despite</i> the merger of the internal market and the area of freedom, security and justice may lead to MS acting unilaterally to protect individuals as they deem fit	Clarify the balance between rights for MS and rights for individuals, and clarify the allocation of legal remedies in a specific MS as well as the locus of accountability for human rights violations
Extraterritorial jurisdiction over crimes in other MS	Relying on this principle undermines mutual trust (since its application requires an investigation into whether the MS deserves to be trusted), and creates differences of opinion between MS and evaluation teams	Limit (or allocate in order to make exclusive) the jurisdiction of each MS to territorial jurisdiction only: this will encourage responsibility over the crimes for which one MS is exclusively competent

3. Domestic factors related to governmental and administrative structures/resources (Management perspective)

Factor	Hypothesis /Explanation	Lessons Learned
Type of legal instrument used for transposition	The use of primary legislation for transposition more often than not results in transposition delays (since it requires parliamentary approval and/or the involvement of advisory boards) ³¹⁴	Avoid, if possible, using primary legislation for transposing EU law, and/or ensure earlier parliamentary involvement in the process
Package deals during the transposition	Governments that often use the package deal approach (transposing several directives at once) are more likely to demonstrate transposition delays	Do not combine the transposition of different directives into a single transposition process that would make the progress of one directive contingent upon that of the other
Gold plating	Gold plating (the adding of extra domestic requirements to the EU standards while transposing) is likely to delay the transposition process, but is expected to promote practical compliance	Gold plating or national ‘icing’ contributes to collective learning
Copy and paste method	Literally transposing EU requirements into domestic legislation is likely to hasten the transposing process, but is expected to have negative effects on practical compliance, since it disregards specific national circumstances	Consider using other (context sensitive) methods of transposition
Suitability of fit	The less the fit between EU requirements and the domestic circumstances, the higher the adaptation costs for Member States, and thus the worse the compliance performance ³¹⁵	Reduce the ambitions of the legislation and/or provide capacity building incentives to support MS

³¹⁴ A striking example is Italy, where the requirement of parliamentary adoption, coupled with the high turnover rate of Italian governments, was, for a long time, the primary source of its compliance problems. See Tallberg, J., Paths to Compliance: Enforcement, Management, and the European Union, International Organization, 2002.

³¹⁵ Goodness of fit can take place at two levels and in two ways: policy and institutional fit. Policy fit refers to the ‘match’ between the text/content of the new EU requirements and the already existing domestic legislation. If the existing national legislation resembles the new EU requirements (that is, a ‘fit’ occurs), compliance is expected to be smooth. Institutional fit refers to the extent to which a country has all the necessary organizational and institutional structures in place to actually comply with the new EU requirements. For example, a country that does not have a proper functioning environmental agency, will show an institutional misfit with new EU requirements that require inspections, and thus this country is not expected to show a smooth compliance record.

'Chinese walls'	The existence of Chinese walls (a lack of communication/coordination between the actors involved in the preparation of legislation and those involved in the execution) negatively influences compliance results	Involve those active in the executive phase in the policy formulation phase, and be aware of the feelings/opinions of the executors on a certain issue (for example, practitioners in the third pillar want less rather than more legislation/changes)
Administrative/institutional capacity	The more resources, personnel, expertise, bureaucratic organization (for example, inter/intra-ministerial coordination) a government has at its disposal, the easier compliance will be ³¹⁶	Provide capacity building via technical assistance, and encourage learning to take place (for example, via peer review)
Veto players	The more veto players in place in a country, the more likely compliance problems will arise	As this factor is particularly important where primary legislation is used for transposition, avoid, if possible, transposing EU directives through primary law and/or involve potential veto players as early as possible
Parliamentary involvement	Parliamentary involvement <i>before</i> the adoption of EU legislation increases the chances of successful compliance	Ensure that national parliaments are involved already from the beginning of the EU legislative process
Economic capabilities	The stronger a country's economic capabilities (measured by its GDP per capita), the more successful its compliance performance	Provide capacity building via funding and subsidies to those MS that lag behind

³¹⁶ One of the most famous examples of the importance of capacity problems is Luxembourg. The main reason why Luxembourg often has problems complying with EU legislation is an administrative overload due to its lack of resources (lack of staff to deal with all legislation coming from Brussels). See Falkner, G., Hartlapp, M., Leiber, S. and Treib, O., Non-Compliance with EU Directives in the Member States: Opposition through the Backdoor?, West European Politics, 2004.

4. Domestic factors related to rationalist / power perspectives (Rationalism)

Factor	Hypothesis /Explanation	Lessons Learned
Governmental position / political priority	EU legislation in line with the governmental position/interests of leading political actors will lead to positive compliance results	Keep in mind that States look at their own interests, and that the MS holding the Presidency (therefore) complies better than other MS
Willingness to comply	Countries with a strong intention to comply (for example, because they are satisfied with the legal instrument as adopted), will show better compliance results	Keep in mind that MS that are not satisfied with the instrument may wait for infringement procedures so as to maintain the old situation for as long as possible
Cost-benefit calculations	When the costs of non-compliance (for example, sanctions, economic losses) outweigh its benefits, countries will be more likely to comply	Increase the (economic/social) costs of non-compliance (non-compliance with criminal legislation does not cost the State anything)
Uploading capacities of a country	The better a country is able to upload its own preferences to the EU level (that is, the more powerful a country is in the negotiations), the more likely it is to comply with the legislative output ³¹⁷	Be aware that different reactions by MS to the EU's legislative output may be (partly) explained by their different capacities to influence the process – the preferences of a specific MS may not be those of others; and keep in mind also that MS with a low capacity to influence the EU policy making process may take 'revenge' by not complying with legislative output.

³¹⁷ Countries that do not manage to upload their own preferences into new EU policy are expected to resort to 'opposition through the backdoor' by not complying with it; (Falkner, G., Hartlapp, M., Leiber, S. and Treib, O., Non-Compliance with EU Directives in the Member States: Opposition through the Backdoor?, West European Politics, 2004). Research shows, however, that this form of opposition does not often occur in practice. An important variant of this hypothesis applies to 'new' Member States. Legislation that was adopted before a country entered the EU results in a situation in which that country did not have the opportunity to influence the legislation. It is thus more likely to create compliance problems than for countries that did have the opportunity to influence the decision making process.

5. Domestic factors related to normative/constructivist perspectives (Constructivism)

Factor	Hypothesis /Explanation	Lessons Learned
Relevance of the issue	If the relevance of a topic (as indicated, for example, by the (media) attention or the political importance attached to a topic) is low in a country, there is less 'pressure' to comply	Try to increase media attention, as this enhances public issue interest, which in turn may enhance public pressure, which in turn enhances systemic issue relevance (for example, for obvious reasons, measures related to terrorism are generally complied with)
Negative public opinion	If public opinion is against a certain policy, a government may be less interested in complying with it	Negative public opinion based on misconceptions of EU policy may be counter-acted if objective media attention is given to the topic at stake
Logic of appropriateness	The longer a country is a member of an organization (for example, the EU), the more this country socializes or internalizes the 'rules of the game' of this organization as its own (socially accepted behaviour), and thus the more likely it becomes that this country will comply	Stimulate processes of socialization (for example, encourage Member States to internalize EU rules as the 'appropriate' behaviour) via seminars, peer review, or through the mobilization of societal actors (for example, think tanks, research institutes, press, NGOs etc.) ³¹⁸

³¹⁸ For example, research leads to more information, which leads to more media attention, which leads to increased interest in an issue, which leads to mobilized public concern, which in turn induces socialization.

Executive Summary

This study focuses on Member State compliance with monitoring, evaluation and inventory mechanisms employed within both the first and the third pillar of the European Union. Following a comprehensive analysis of existing practice, it seeks to offer practical recommendations on how to improve compliance with such mechanisms within the third pillar.

At first glance, the existing evaluation mechanisms are predominantly of an administrative nature. Their general feature is that they are fragmentary, incoherent and inconsistent. In other words, the current mechanisms are both insufficient and inefficient, in the sense that they do not guarantee implementation of the common binding norms of the third pillar.

One of the foundations or preconditions of the principle of mutual recognition is that the Member States share common ground regarding fundamental principles of their legal systems, such as the rule of law, the protection of human rights, the principle of legality, the principle of ‘subsidiarity’, the principle of proportionality, and other principles common to all criminal justice systems. With this in mind, how can the Member States cooperate effectively if full mutual trust in the legal system of another Member State, or its implementation, is not present?

In order to be able to further promote the necessary mutual trust within these areas, it is imperative to examine and strengthen the functioning of monitoring, evaluation and inventory mechanisms existing within the third pillar of the European Union. For us to produce appropriate recommendations on how to improve the respective mechanisms and compliance with these mechanisms by the Member States and, eventually, to strengthen mutual trust within the European Union, a careful examination of the content and process related factors that at the same time both hinder and contribute to compliance with third pillar mechanisms is required. Such an examination necessitates a study of the extent to which the degree of compliance is contingent upon the actual organizational structure (approach, working method, constellation of actors involved) within which these monitoring mechanisms operate, as well as the way in which these additional factors may also hinder or contribute to compliance.

Because any proposals to improve actual compliance with the monitoring mechanisms within the domain of criminal matters may also be derived from an examination of the experiences with such mechanisms in the first pillar, our recommendations are based on a study on the functioning of monitoring, evaluation and inventory mechanisms in both the third pillar and the first pillar.

The study was structured according to the following research questions:

4. Which, content and process related factors hinder (better) compliance with monitoring, evaluation and inventory mechanisms within both the first and third pillars of the EU?
5. To what extent does the degree of compliance depend on the approach and working methods used, and the constellation of actors involved in these monitoring, evaluation and inventory mechanisms?

6. What adjustments to monitoring, evaluation and inventory mechanisms may improve the follow-up to the outcomes of such mechanisms by the Member States?

The relevance of these research questions can be seen from the current absence of an effective monitoring system. When ratified, the Treaty on the Functioning of the European Union will provide for a formal framework of such an implementation scheme (Article 70 TFEU). However, given the current uncertainties as to when the Treaty will be ratified, the transition period of five years after the entering into force of the Treaty of Lisbon regarding third pillar legislation, and the increasing number of situations in which legislation is not applicable to all Member States, it is important to begin now with the process of developing better monitoring systems.

Based upon existing practices, this report analyses those factors that hinder (better) compliance (and follow up) with monitoring and evaluation mechanisms in the first and third pillars of the European Union. We base our findings on an in depth analysis of existing literature on this topic (see the Annex to this report for a schematic overview of compliance factors, as well as chapters 2 for detailed descriptions), coupled with evidence based analysis of real life cases and experiences with such monitoring and evaluation mechanisms (see chapters 4 and 5). This enables us to draw from the experiences with monitoring and evaluation mechanisms in a wide variety of situations, such as in the OECD, the first pillar and the third pillar of the European Union (see, in particular, chapters 3, 4 and 5).

The main – and very obvious – lesson to be derived from these analyses is the realization that a well functioning monitoring or evaluation mechanism needs a context sensitive approach, not a one size fits all or implementation biased approach. A context sensitive approach needs to take account of the following crucial elements:

4. Monitoring and evaluation mechanisms to a large extent stand or fall with the degree to which its members learn from each other and thereby adhere to the rules, which in turn depends upon the level of commitment of its members. For every monitoring or evaluation mechanism that is to be established at the European Union level, we carefully need to examine the political will amongst Member States to achieve the required goal. Without a basic level of trust and commitment among the members, even the best designed evaluation or monitoring mechanism is subject to failure.
5. A second necessary element in order to allow for a context sensitive approach is the realization that specific Member State factors do matter and that these domestic factors are, generally speaking, rather ‘sticky’ or ‘culturally determined mental patterns’, and thus difficult to overcome. Actors that play a dominant role in ensuring follow up in these mechanisms need to be(come) sensitive to these domestic specificities and need to have the freedom, and be able, to react upon them.
6. A third element which particularly comes to the fore in the field of criminal law, is that the interdependence of Member States and exclusive nature of the legislation, stimulate compliance at all levels, both in relation to compliance with legislation, as well as with the outcomes of evaluation findings. A spirit supportive of not only the instrument, but also its European Union nature and cooperation with other Member States, will stimulate

compliance. This could also be promoted by a common policy on certain areas, which is not only formally or symbolically supported by the Member States, but also appears from what they actually do.

The study identifies three possible scenarios:

4. Monitoring and evaluation mechanisms based on peer review

On a positive note, monitoring and evaluation mechanisms that largely rest on peer review among the member countries they may be beneficial in stimulating mutual learning via peer pressure. Member States seem to be relatively open to accepting peer review from other Member States, and the intensive, non-confrontational and non-legalistic nature of the interaction in most peer review systems generally leads to an ongoing and informal information-sharing experience from which all partners may benefit.

5. Monitoring and evaluation mechanisms based on independent, ‘expertocratic’ reporting

It is argued that, only if reporting is done in an independent, uniform, objective and standardized way, is it possible to arrive at a genuine comparability of national policies and of their respective impacts. Independent experts might be better placed to report accurately and truthfully on the Member States’ performances.

6. Monitoring and evaluation mechanisms based on sanctioning

The third scenario incorporates the use of sanctions to stimulate (by imposing fines) or to force (through the launch of infringement proceedings) Member States to comply. However, while the presence of clear sanctioning mechanisms is a positive incentive for Member States to comply, we require the necessary political commitment to make the mechanism work in practice.

Without necessarily indicating a preference for a particular scenario or a specific combination of scenarios, we would emphasize that any system that intends to promote compliance with the implementation of European Union agreements, as well as compliance with monitoring and evaluation results, needs to adhere to a multi-step approach. Without precisely specifying the number of steps needed, an analysis of compliance outcomes in the first and the third pillars of the European Union seems to at least indicate the need for a first inventory step, where information is collected in the different Member States in the most uniform and objective manner. This process of data collection should be undertaken according to the following conditions:

Data collection is most likely to result in reliable and comparative information, which is crucial for checking compliance and stimulating follow up, when:

- the monitoring or evaluation process is institutionalized (Member States are more likely to participate in, and accept the outcomes of, institutionalized mechanisms, since institutionalisation creates a dynamic of its own);

- the monitoring or evaluation process takes place on the basis of an iterative process supported by a permanent structure (if Member States are aware that their performance is judged regularly and repetitively, the probability that they will feel compelled to continue paying attention to the issue(s) discussed is greater than where ad hoc evaluations are concerned, due to the follow up element that it implies);
- the monitoring or evaluation process uses common European, precise indicators for assessment (it is self-evident that progress can only be measured comparatively if there are precisely defined and explicit indicators that limit the leeway for interpretation for those who report on national progress);
- the monitoring or evaluation process results in prescriptive and express policy recommendations that allow for effective monitoring and evaluation, and that give the process a forward-looking orientation. Only with such recommendations can the evaluation process be constructive and enhance compliance, while preventing the emergence of evaluation fatigue.

In order to ensure actual compliance with monitoring and evaluation outcomes, we would, however, support the shift towards a more flexible approach reaching beyond an initial uniform and generally applied methodology; an idea based on our general conclusion that well-functioning monitoring and evaluation mechanisms need a context sensitive approach that takes into consideration political will and commitment, as well as domestic circumstances.

Leesvervangende samenvatting

Dit onderzoek richt zich op ‘compliance’ met monitoring- en evaluatiemechanismen in de eerste en derde pijler van de EU en presenteert praktische aanbevelingen met het oog op het verbeteren van ‘compliance’ met dergelijke mechanismen in de derde pijler.

Op het eerste gezicht lijken de bestaande evaluatiemechanismen in de derde pijler voornamelijk van beschrijvende aard. Over het algemeen zijn ze fragmentarisch, incoherent en inconsequent. Kort gezegd geldt, dat de huidige evaluatiemechanismen onvoldoende en inefficiënt zijn in de zin dat ze implementatie met algemeen bindende normen in de derde pijler niet kunnen garanderen. In het licht van het beleden wederzijdse vertrouwen tussen de lidstaten van de EU is dat merkwaardig.

Een van de uitgangspunten van het beginsel van wederzijdse erkenning is dat de lidstaten bepaalde fundamentele beginselen betreffende het rechtstelsel gemeen hebben. Voorbeelden van dergelijke beginselen zijn rechtstatelijkheid, de bescherming van mensenrechten, het legaliteitsbeginsel, subsidiariteit, proportionaliteit en ander principes die van toepassing zijn op alle strafrechtstelsels. Echter, hoe kunnen we verwachten dat lidstaten samenwerken als het wederzijds vertrouwen ontbreekt in de strafrechtelijk systemen van andere lidstaten, of als de implementatie van Europese beginselen in deze systemen niet is voltooid?

Met het oog op het versterken van dit wederzijds vertrouwen op het gebied van strafrecht is het bestuderen en vervolgens verbeteren van het functioneren van de huidige monitoring- en evaluatiemechanismen in de derde pijler een logische stap. Teneinde bruikbare aanbevelingen te kunnen geven om de bestaande mechanismen én de ‘compliance’ van lidstaten met de uitkomsten daarvan te verbeteren is een grondig onderzoek nodig dat focust op het identificeren van inhoud- en procesgerelateerde factoren die ‘compliance’ met derde-pijler evaluatiemechanismen (ver)hinderen dan wel stimuleren. Een dergelijk onderzoek vereist dat men het ontwerp en de organisatie (bijvoorbeeld de gekozen benadering en werkmethode alsook de betrokken actoren) van dergelijke monitoring- en evaluatiemechanismen bekijkt, evenals de manier waarop deze specifieke manier van organiseren de mate van ‘compliance’ door lidstaten beïnvloedt.

Het is vanzelfsprekend, dat het bestuderen van ervaringen met monitoring- en evaluatiemechanismen in de eerste pijler van de EU erg nuttig kan zijn gezien de algemene doelstelling van dit onderzoeksproject. De aanbevelingen ter verbetering van ‘compliance’ met monitoring- en evaluatiemechanismen in de derde pijler zijn dan ook gestoeld op een analyse betreffende het functioneren van dergelijke mechanismen in zowel de derde als de eerste EU pijler.

Aan het onderzoek liggen de volgende onderzoeksvragen ten grondslag:

1. Welke inhoudelijke en procesmatige factoren staan in de weg aan een betere compliance van de lidstaten met EU eerste pijler en EU derde pijler monitoring-, evaluatie- en inventarisatiemechanismen?
2. In hoeverre hangt de mate van compliance af van de aanpak, werkwijze en samenstelling van beoordelingsgremia van deze mechanismen?

3. Welke aanpassingen aan deze mechanismen zouden de follow-up door de lidstaten kunnen verbeteren?

De relevantie van deze onderzoeksvragen ligt in de huidige afwezigheid van een effectief monitoring- en evaluatiesysteem. Na ratificatie zal het Verdrag betreffende de Werking van de Europese Unie een formeel kader voor een dergelijk ‘implementatieschema’ bieden (zie artikel 70 VWEU). Echter, vanwege de huidige onzekerheid betreffende het tijdstip van ratificatie van dit Verdrag, de overgangperiode van vijf jaar na het in werking treden van het Verdrag van Lissabon wat betreft de wetgeving betreffende de derde pijler, én het toenemende aantal situaties waarin wetgeving niet op alle lidstaten van toepassing is, is het goed om nu al te beginnen met het ontwikkelen van betere monitoring- en evaluatiesystemen.

Zoals gezegd treft men in dit rapport een analyse aan betreffende de factoren die (betere) ‘compliance’ met bestaande monitoring- en evaluatiemechanismen in de eerste en derde pijler (ver)hinderen. We baseren onze bevindingen op een grondige analyse van de bestaande academische literatuur betreffende het thema ‘compliance’ (zie Annex voor een schematisch overzicht, en zie hoofdstuk 2 voor beschrijvingen) en op een meer empirische analyse van ervaringen met de huidige monitoring- en evaluatiemechanismen (zie hoofdstukken 4 en 5). Op deze manier kunnen we leren van ervaringen met dergelijke mechanismen in een contextuele diversiteit, die naast ervaringen in verscheidene beleidsgebieden in de eerste pijler en de derde pijler van de EU ook een deel van de evaluatiepraktijk in de OESO omvat (zie de hoofdstukken 3, 4 en 5).

De belangrijkste – en overduidelijke – les die we trekken uit onze analyse is het feit dat voor het adequaat functioneren van monitoring- en evaluatiemechanismen, een ‘contextgevoelige benadering’ noodzakelijk is, en dat een ‘one-size-fits-all benadering’ of een benadering die slechts op het evalueren van implementatie gericht is niet voldoet. Een benadering waarbij men de bestaande diversiteit in de EU alsook het bredere wetgevingskader waarin implementatie moet plaatsvinden in ogenschouw neemt zou, specifiek, rekening moeten houden met de volgende cruciale aspecten:

1. De effectiviteit van monitoring- en evaluatiemechanismen staat of valt grotendeels met de mate waarin lidstaten wederzijds leren van ervaringen, en hangt eveneens af van de mate waarin lidstaten zich in het algemeen aan regels houden. Wat beide betreft geldt dat het commitment van de lidstaten cruciaal is. Dit betekent dat we voor elk monitoring- en evaluatiemechanisme dat op EU-niveau geïntroduceerd is/wordt moeten onderzoeken of het commitment om het gestelde doel te behalen überhaupt aanwezig is. Zoals eerder gesuggereerd is de mate van commitment sterk afhankelijk van de mate van vertrouwen in het functioneren van de strafrechtelijke systemen van andere lidstaten. Hierop volgend geldt dat zonder een bepaald niveau van vertrouwen onder de lidstaten zelfs het best onderworpen monitoring- en evaluatiemechanisme gedoemd is te mislukken.
2. Een tweede element dat we in overweging moeten nemen betreft het feit dat factoren op het niveau van de lidstaten zelf belangrijk zijn voor de mate waarin monitoring- en evaluatiemechanismen effectief functioneren, en dat deze lidstaatspecifieke factoren over het algemeen cultuurbepaalde, moeilijk te veranderen patronen betreffen. Actoren die een belangrijke rol spelen in de follow-up fase van de monitoring- en evaluatiemechanismen

zouden zich bewust moeten zijn van de lidstaatspecifieke factoren die een rol spelen in het al dan niet (kunnen) effectueren van ‘compliance’. Eveneens zouden zij de ruimte én capaciteiten moeten krijgen om hierop in te kunnen spelen.

3. Een derde element, dat specifiek op de voorgrond treedt op het gebied van strafrecht, is het idee dat wederzijdse afhankelijkheid cq verbondenheid van lidstaten en de exclusiviteit van de te implementeren wetgeving een positieve invloed op de mate van ‘compliance’ hebben, zowel betreffende ‘compliance’ met wetgeving alsook ‘compliance’ met de uitkomsten van evaluatiebevindingen. Het spreekt voor zich dat een insteek die niet alleen de betreffende wetgeving, maar ook haar Europese aard én de samenwerking met andere lidstaten ondersteunt ‘compliance’ bevordert. Het formuleren van een algemeen Europees beleid op zekere gebieden kan een dergelijke insteek bevorderen en ertoe leiden dat EU wetgeving niet alleen formeel of symbolisch, maar ook in de praktijk door lidstaten ondersteund wordt.

Concreet gezien identificeren we de volgende scenario’s als het gaat om het ontwerp van monitoring- en evaluatiemechanismen:

1. Monitoring- en evaluatiemechanismen gebaseerd op ‘peer review’

Monitoring- en evaluatiemechanismen die grotendeels op ‘peer review’ gebaseerd zijn, kunnen positief zijn in de zin dat zij, door de aanwezige ‘peer pressure’, wederzijds leren stimuleren. Lidstaten lijken relatief ontvankelijk te zijn voor ‘peer review’ door andere lidstaten; de intensieve, niet op confrontatie gerichte aard van de interactie in de meeste ‘peer review’ systemen kan in het algemeen leiden tot een continue en informele informatie-uitwisseling waarvan elke deelnemer kan profiteren.

2. Monitoring- en evaluatiemechanismen gebaseerd op rapportering door onafhankelijke experts

Het lijkt duidelijk dat rapporteren op een onafhankelijke, uniforme, objectieve en gestandaardiseerde manier de noodzakelijke basis is om tot een werkelijke vergelijkbaarheid van nationaal beleid en van de impact dat dit nationaal beleid heeft te komen. Het ligt voor de hand dat onafhankelijke experts zich inderdaad in een relatief goede positie bevinden om accuraat en waarheidsgetrouw te kunnen rapporteren betreffende de prestaties van de lidstaten.

3. Monitoring- en evaluatiemechanismen gebaseerd op sanctionering

Het derde scenario houdt het gebruik van sancties in om ‘compliance’ door lidstaten te stimuleren (door oplegging van boetes) of letterlijk af te dwingen (door het starten van infractieprocedures). Echter, hoewel de aanwezigheid van een duidelijk sanctioneringmechanisme inderdaad een positieve stimulans kan zijn om ‘compliance’ door lidstaten te effectueren, vormt de politieke wil om het mechanisme ook echt te laten functioneren een noodzakelijke voorwaarde voor haar effectiviteit.

Zonder een voorkeur voor enig scenario of een specifieke combinatie van scenario’s te willen uitspreken, willen we tenminste één ding wel duidelijk aangeven: elk systeem dat het stimuleren van ‘compliance’ met Europese afspraken (in de brede zin van het woord) alsook het stimuleren van ‘compliance’ met de uitkomsten met monitoring- en evaluatiemechanismen beoogt, zou gekenmerkt moeten zijn door een uit meerdere stappen bestaande benadering. Zonder een definitief aantal stappen te willen suggereren als noodzakelijk, willen we er op wijzen dat de resultaten uit de analyse zoals gepresenteerd in dit rapport lijken aan te geven dat er in elk geval behoefte is aan een inventariserende stap waarin informatie in de verschillende lidstaten

verzameld wordt op basis van de grootst mogelijke uniformiteit en objectiviteit. Voor een dergelijk proces van informatieverzameling gelden de volgende voorwaarden:

De kansen dat het verzamelen van gegevens resulteert in betrouwbare en vergelijkbare informatie, cruciaal voor het controleren van 'compliance' en het stimuleren van follow-up, zijn het grootst als:

- het monitoring- of evaluatieproces geïnstitutionaliseerd is (institutionalisering stimuleert de deelname van lidstaten en hun acceptatie van evaluatieresultaten; institutionalisering genereert een eigen dynamiek);
- het monitoring- of evaluatieproces continu is en ondersteund wordt door een permanente structuur (de kans dat lidstaten zich gedwongen voelen aandacht te blijven geven aan onderwerpen die ter discussie gesteld worden is groter als zij zich ervan bewust zijn dat hun prestaties op basis van regelmaat en herhaling, in plaats van op een *ad hoc* basis, geëvalueerd worden, omdat een permanente manier van evaluatie impliceert dat de follow-up door lidstaten gecontroleerd wordt);
- het monitoring- of evaluatieproces algemeen Europese en precieze indicatoren voor evaluatie inhoudt (het is duidelijk dat voortgang alleen op vergelijkbare wijze gemeten kan worden als er precieze en expliciete indicatoren zijn, die de ruimte voor interpretatie door rapporteurs én geëvalueerden beperken);
- het monitoring- of evaluatieproces resulteert in prescriptieve en expliciete beleidsaanbevelingen (alleen in de aanwezigheid van aanbevelingen die het mogelijk maken om voortgang effectief te monitoren en evalueren en die het proces een op de toekomst gerichte oriëntatie geven kan het evaluatieproces constructief en 'compliance-bevorderend' zijn, en wordt 'evaluatiemoetheid' voorkomen).

Om 'compliance' met de uitkomsten van monitoring- en evaluatieprocessen ook werkelijk in de praktijk te bereiken, stellen we voor een meer flexibele benadering toe te passen, die buiten deze eerste uniforme en algemeen toegepaste benadering omgaat. De oorsprong van dit voorstel gaat terug op onze algemene conclusie dat teneinde werkelijk adequaat functionerende monitoring- en evaluatiemechanismen te formuleren, een contextgevoelige benadering, die zowel politieke commitment als lidstaatspecifieke omstandigheden in overweging neemt, een noodzakelijke voorwaarde is.