



A **POLICY ENGINE** FOR EVIDENCE-BASED INSTITUTIONAL CHANGE



PRACTITIONER'S GUIDE

Generating evidence, developing policy recommendations
and tracking change in evolving criminal justice systems



This publication was produced with the financial support of the European Union. Its contents are the sole responsibility of the OSCE and do not necessarily reflect the views of the European Union.

The contents of this publication, the views, opinions, findings, interpretations and conclusions expressed herein are those of the authors and contributors and do not necessarily reflect the official policy or position of the OSCE and its participating States. Although the OSCE has invested the utmost care in the development of this publication, it accepts no liability for the accuracy and completeness of the information, instructions and advice provided, nor for misprints. For these reasons, no claims can be made against the OSCE in respect of potential consequences which may arise out of information or conclusions contained in this publication.



REGIONAL TRIAL MONITORING PROJECT
COMBATting ORGANIZED CRIME AND CORRUPTION
IN THE WESTERN BALKANS

Author:

Eirini Patsea (Senior Programme Co-ordinator, OSCE/Conflict Prevention Centre, Policy Support Section)

Contributing legal experts:

Rebecca Agule, Francesco de Sanctis

Editorial and information design:

Kristina Allram-Naaijer

osce.org/westernbalkanstrialmonitoring

Conflict Prevention Centre, OSCE Secretariat
Wallnerstrasse 6
A-1010 Vienna, Austria
pm-cpc@osce.org osce.org/cpc

Contents

Abbreviations	6
Glossary	7
About this guide	15
Introduction	19

The Steps

Scope	26
Collect	50
Analyse	73
Explain	110
Recommend	138
Track	158
Closing the Loop	176
Annex I- Deep Dives	178
Annex II- Further Reading	181

Key Abbreviations

C

CPC

Conflict Prevention Centre, OSCE Secretariat

E

EBTM

Effectiveness-Based Trial Monitoring

ECtHR

European Court of Human Rights

EU

European Union

M

MC

OSCE Ministerial Council

O

OCC

Organized crime and corruption — the case typology used as the Guide's worked application

ODIHR

OSCE Office for Democratic Institutions and Human Rights

OSCE

Organization for Security and Co-operation in Europe

R

RBM

Results-Based Management

RTM

The OSCE Regional Trial Monitoring Project

Glossary



A

Actor-and-authority map

The record of which institution or office holds the legal mandate, decision authority and implementation responsibility for the practice in question. It establishes who can lawfully change the practice; it does not assign blame for it.

Applicable standard

The legal, procedural, human-rights or policy rule or requirement against which observed practice is assessed. Identified precisely enough to show why the practice matters and calibrated to the applicable jurisdiction and context.

Assumption

A condition that must hold for the design to work but that the practitioner does not control — in Scope, about case flow, access, publicity of hearings or classification; in Recommend and Track, about the reform pathway. Assumptions are recorded, tested and as needed, acted upon not to become risks.

B

Baseline (context baseline)

What is already known, believed or claimed about the practice before monitoring begins, together with what remains uncertain. It frames the reform question and supplies the assumptions monitoring must test; it is not itself evidence of current practice. Distinct from Tracking baseline.

Bounded inference

The discipline that a conclusion must not extend beyond what the evidence base, monitoring design, comparison, scope and uncertainty can responsibly support.

C

Capacity

Assesses whether justice-system actors and institutions have and apply the skills, routines, legal and procedural tools, reasoning practices, resources, support and coordination needed to perform the relevant function. A practice- or institution-level assessment, never a personal rating. In external communication it is often expressed as the operational conditions for effective case handling.

Closed loop

A reform process in which implementation evidence, new findings and subsequent learning can return to earlier work and inform ongoing or new processes.

Coding

The consistent assignment of recorded facts to predefined categories and variables so that material from different cases, courts and periods can be compared. Coding makes comparison possible; it does not, by itself, make a claim.

Comparability

The requirement that anything compared belongs to a like group — offence type, defendant status, procedural route, court level, complexity, period — and that the denominator is stated. Without it, an apparent difference between cases may be an artefact of how they were selected.

Concern (and good practice)

The paired categories in which observed practice is recorded. A concern is an observable feature that may be inconsistent with an applicable standard or expected practice and warrants further examination; a good practice is one that supports legality, fairness, quality or reliable functioning. Both are recorded; neither is yet a finding.

Confidence

The extent to which the available evidence supports a material claim.

Confidence language

The controlled vocabulary for stating evidentiary strength: the evidence shows; the evidence indicates; the evidence suggests; the evidence raises a question about; the evidence does not allow a conclusion on. Applied consistently across every product. Not a stylistic preference, and not a taxonomy that must be assigned to every finding.

Contribution

The claim this Guide's evidence can ordinarily support: that a institutional performance condition plausibly contributes to an observed pattern. Weaker than causation, and it does not establish that the condition alone produced the pattern or that the pattern would disappear without it.

D

Data

Recorded or structured information – for example, coded observations, categories, counts or statistics. Data do not, by themselves, establish a claim; they become evidence for a claim only once their source, relevance, reliability and limitations are assessed. Not a synonym for evidence

Deep Dive

A focused inquiry initiated when the ordinary work of a Step leaves a material question unresolved and additional evidence or analysis is likely to affect the next decision. Returns its results to the Step it was intended to strengthen; it is not a separate reform track.

Denominator

The defined set against which a count is expressed — all monitored hearings, all eligible cases, all cases followed under one modality. Stating the denominator is what makes a proportion meaningful. Modality and coverage change the denominator, and figures drawn from different denominators are not comparable.

Dependency

An external action, resource, institution, sequence or condition on which implementation or the expected result materially relies. Every dependency contains an assumption — that what is relied on will arrive; not every assumption names a dependency. Assumptions are tested; dependencies are sequenced, owned, and where blocked reported as bottlenecks.

Dimensions of effectiveness

The formal term for Capacity, Fairness and Efficiency. “Lens” may be used as informal explanatory language for how a dimension is applied but is not a second formal label.

Driver / contributing condition

A legal, institutional, operational, professional or external condition, supported by evidence, that plausibly contributes to the occurrence, recurrence or persistence of a pattern. Not a sole, necessary or proven cause.

E

Effectiveness

In EBTM, the demonstrated ability of justice-system actors and institutions to fulfil their mandates in practice, assessed through Capacity, Fairness and Efficiency rather than inferred from conviction rates, case resolution or speed alone.

Effectiveness-Based Trial Monitoring (EBTM)

The Guide's methodology for using systematic trial-monitoring and related evidence to assess justice-system effectiveness in observable practice, and to connect that evidence to analysis, explanation, recommendation design and tracking.

Efficiency

Assesses whether time, sequence, complexity, process and resources are managed without avoidable delay, repetition or waste, and without sacrificing fairness, legality or quality.

Eligibility rule

The written rule determining whether a case falls inside the field of inquiry, applied consistently and settled in advance, together with the stated treatment of borderline cases. Substantive similarity to the subject matter does not make a case eligible where the formal criteria are not met.

Eligible field

The boundary of the inquiry: the subject matter, institutional site, jurisdiction, geographic and temporal coverage, unit of observation and intended unit or units of analysis. It defines what qualifies for monitoring and, by exclusion, what does not. A definition of criteria rather than a set of cases; the cases selected from within it are the monitored pool. It is the reference point against which the monitored pool's coverage and limits are stated.

Enabling condition

A condition that must exist for the practice targeted by the main recommendation to become possible, likely, safe and sustainable. Sub-recommendations exist to create or protect enabling conditions; they are not a list of desirable extras.

End-of-Case Assessment

A structured case-level assessment conducted at the defined endpoint of a monitored case.

Evidence

Source material assessed as relevant and sufficiently reliable for a specific claim, and used to support, qualify or challenge it.

Evidence base

The body of source material and data available to assess and address the reform question and to support a given claim, including documented gaps, limitations and source distinctions.

Evidence strategy

The written, version-controlled output of Scope and the instruction to Collect. It records the reform question, baseline, subject-matter boundary, selection protocol, monitored pool, observation window, monitoring modality, source requirements, access plan, collection variables, evidence gaps and review points — and the reasons for each, so that another practitioner could reconstruct every decision.

Explanatory account

A bounded, evidence-based account of the condition or combination of conditions that plausibly contributes to an observed pattern, the pathway through which it may operate, and the limits of that conclusion. Does not necessarily establish strict causation.

Explanatory status

The formal categorical statement in Explain of how far the evidence supports a proposition: strongly supported driver, plausible driver, working hypothesis, or evidence gap.

F

Fairness

Assesses whether justice-system practice respects applicable fair-trial guarantees and related requirements – equality before the law, equality of arms where applicable, impartiality, independence and other relevant safeguards. Not a loose synonym for public legitimacy.

Finding

A bounded conclusion reached through Analyse about what the evidence shows in observable practice against a stated standard, within explicit evidentiary and inferential limits.

Full monitoring

The trial monitoring modality combining direct observation of hearings, continuous case and review of the relevant legal acts, closing with an End-of-Case Assessment.

G

Good practice

A specific, source-linked professional or institutional practice that supports legality, fairness, quality, efficiency or reliable justice-system functioning, goes beyond meeting the legal minimum and may offer transferable learning/ Compliance is not good practice. It need to be tested with the same evidentiary discipline as a concern — source, recurrence, applicable standard, stated limits, and whether it survives the departure of the individual who performed it. A practice clearing both is a candidate for preservation, scaling or standardization.

I

Implementation stage descriptor

A qualitative label used in Track to describe the nuanced stages of reform implementation.

Implementation stage narrative

The written account that gives the implementation stage descriptors their meaning: what moved and what stalled, which milestones and evidence were decisive, whether the anticipated pathway is holding, what contrary evidence remains, and what would change the judgement at the next review.

In-court performance

Observable conduct of proceedings in court – including hearing management, party participation, evidentiary presentation and procedural safeguards – assessed as institutional or professional practice, not as a personality or behavioural rating of individual actors.

Institutionalisation

The condition in which a verified practice change is sustained and supported by ordinary institutional arrangements, so that it is reasonably likely to endure without continued external attention..

Issue (case-level issue)

A concern or good practice that appears more than once, or with material effect, within a single case. The issue is the unit that carries analysis from the hearing level to the case level; recurrence of the same issue across comparable cases is what produces a cross-case pattern.

L

Legal act / judicial act / prosecutorial act

Legal act is the umbrella term. A judicial act is a court-issued judgement, decision, ruling or order. A prosecutorial act is a prosecution-issued document, including an indictment.

Level of change

The level at which the main recommendation is pitched: changed institutional performance or professional practice where the diagnosis and reform maturity support it, or an output or process where commitment is weak, evidence is still emerging or the pathway is not yet operationally clear. Not a hierarchy of ambition.

Level of claim

The level at which the evidence supports drawing a conclusion – a single hearing or legal act, an individual case, a cross-case pattern, or a systemic practice within a defined scope.

M

Milestone

A predefined, evidence-based marker of meaningful progress used to make movement relevant to Action, Uptake and Change observable.

Monitored case

A case from the monitored pool that has been confirmed, has entered its observation window and is being actively recorded.

Monitored pool

The subset of eligible cases or proceedings selected for monitoring under the approved selection methodology. Because it is purposively selected, the pool is designed to reveal practice, not to represent the wider eligible caseload.

P

Pattern (cross-case pattern)

Recurrence of the same issue or practice across comparable cases, courts, actors or periods, established by evidence rather than impression. A pattern is not yet an explanation, and recurrence alone does not make it systemic.

Performance condition

A condition in the working environment that affects whether actors can and are likely to perform the desired practice – for example, information and feedback, tools and resources, incentives, knowledge and skills, assignment or institutional capacity, or professional norms.

Policy cycle

The broader, recurring process through which a policy problem is examined, responses are designed and implemented, results are reviewed, and subsequent action is adjusted. The Guide's policy engine is the specific method used within that broader cycle.

Policy engine

This Guide's six-step evidence-to-reform method: Scope → Collect → Analyse → Explain → Recommend → Track.

Practitioner

The single reader this Guide addresses, standing in for anyone applying the method — legal expert, policy specialist, monitor, analyst, adviser or programme manager — and in practice often a team or an institution. A drafting device for giving clear instructions, not an assumption about who does the work.

Procedural event

A specific procedural action, occurrence or decision observable in a hearing or legal act and potentially relevant to the reform question. It becomes analytically significant only through its context, recurrence or relationship to other evidence.

Productive hearing / hearing productivity

A hearing is productive where it advances the case procedurally or substantively — taking evidence, hearing a witness, resolving a motion, issuing a ruling, clarifying the indictment or preparing the next stage. Attendance or the passage of time does not make a hearing productive. Hearing productivity is the corresponding measure across a case or pool.

Proportionality

Used in context. In legal analysis, proportionality asks whether a restriction, measure or sanction is justified and not excessive under the applicable standard. In methodological or design decisions, it asks whether the intensity, cost or risk of an inquiry or response is justified by its decision value and stakes.

Purposive selection (purposive prioritization)

Deliberate, non-random selection of eligible cases according to specified, recorded criteria indicating diagnostic value for the reform question. It is not probability sampling: a purposive pool may show that a pattern recurs in the monitored material but does not by itself establish prevalence across the wider caseload.

Recommendation

A reasoned proposal directed to one or more competent actors to change, establish, protect, discontinue or test a policy, rule, institutional arrangement, process or professional routine, to support a defined improvement in practice.

Recommendation target

The specific change in practice or condition that a recommendation unit seeks to bring about and that the evidence supports. Narrower than the reform objective, which explains why the target matters. Track assesses the target, not the objective.

Recommendation unit

The smallest coherent block of reform design within the broader reform recommendation package, ordinarily consisting of one main recommendation together with the sub-recommendations required to create its enabling conditions.

Reform objective

The broader improvement in justice system functioning to which a recommendation unit is intended to contribute, for example greater legal certainty, more consistent rights protection, or more durable justice outcomes. It supplies direction and purpose; it is not what Track measures.

Reform question

A neutral, answerable question about institutional practice that defines what the monitoring exercise is intended to make visible, without presupposing a defect, cause or responsible actor.

Reform recommendation package

The selected set of recommendation units through which the overall theory of change is put into practice, showing how the units relate, depend on and sequence with one another.

Relitigation

Re-examining whether an individual decision was substantively right. EBTM never does this. It assesses observable practice against stated standards and is not a form of appeal, judicial review or second-instance assessment — a boundary that protects both judicial independence and the credibility of the monitoring.

Results chain

The expected sequence linking enabling activities or processes to outputs or outcomes as changed conditions, changed institutional or professional practice, and to contribution to the broader reform objective. The results chain simply makes explicit how one level of change is expected to contribute to the next, while recognizing that several pathways may operate at the same time and that the sequence may need to be revised as implementation unfolds.

Results-Based Management (RBM)

An approach that makes explicit the relationships among activities or processes, outputs, intended changes and longer-term results. RBM logic helps structure reform design and tracking; it does not, by itself, establish that an intervention caused an observed result.

Routing

The recorded decision at the end of a step about where the work goes next: forward to the next step, back for targeted collection or analysis, into a Deep Dive, or to a decision not to proceed. Progression is a decision that must be justified, not a default.

S

Safeguard

Used in two senses. A legal or procedural safeguard is a protection built into the justice process — a fair-trial guarantee, reasoning requirement or review right. An analytical safeguard is a discipline applied to the practitioner's own work — neutral outcome language, testing rivals, triangulation, preserving uncertainty. Both are stated explicitly; neither is assumed.

Scope of generalization

The cases, institutions, practices, periods, jurisdictions or other population to which a finding may responsibly be extended. Stated explicitly whenever a claim moves beyond the directly observed unit.

Source provenance

The information needed to establish where source material came from and how it was obtained: producer or collector, date, method, version, access conditions and relevant limitations.

Sub-recommendation

A supporting recommendation within a unit that creates or protects an enabling condition required for the main recommendation to work. Functionally different from the main recommendation, not merely smaller or less important, and not a free-standing task.

System stability account

A bounded inquiry into why an established practice may persist despite formal reform or institutional attention, examining the functions, risks, incentives, dependencies, norms and operating conditions that may make the existing routine easier or more stable than an alternative. It is the mirror of the theory of change: the theory of change sets out what must shift for a different practice to become possible; the system stability account sets out what currently holds the existing one in place. Produced in Explain and carried into Recommend, so that reform is designed against the conditions reproducing the practice rather than layered on top of them.

Systemic

A recurring practice that the evidence shows is embedded in, governed by or traceable to a common rule, workflow, routine or institutional arrangement within a defined institutional or procedural scope. Systemic does not automatically mean that the practice occurs throughout the entire domestic justice system or establishes population prevalence.

Systemic finding

A finding that a cross-case pattern is embedded in, governed by or traceable to a common rule, workflow, routine or institutional arrangement within a defined scope. It requires evidence of the common arrangement itself, not only evidence of recurrence.

T

Theory of change

The reasoned account of how changing specified, diagnosed conditions or mechanisms is expected to alter institutional or professional practice and contribute to the broader reform objective, including the assumptions, dependencies and risks on which the pathway relies.

Tier 2 monitoring

An RTM Project term for the document-based modality relying on authorized written sources — indictments, procedural decisions, judgements, appeal and retrial decisions, official transcripts or hearing records, public registers and statistics — without systematic observation of hearings. It broadens documentary coverage and is a controlled complement to direct observation; it cannot establish conduct in court.

Tracking baseline

The bounded starting position against which Track assesses implementation progress for a recommendation unit. It may be quantitative or qualitative, but it should correspond to the recommendation target and be supported by evidence appropriate to that starting claim and be recorded before tracking evidence is collected, so that the comparison is not reconstructed with hindsight. Distinct from the context baseline in Scope, which describes the field before monitoring begins, may carry contested or unverified material, and is revised as understanding improves.

Triangulation

Testing a claim using independent or meaningfully complementary sources. Repeated accounts across several channels are not independent corroboration.

V

Verification

The process of checking a material fact, source or claim against reliable underlying records or other evidence appropriate to the proposition under test.

ABOUT THIS GUIDE

I. Purpose, scope and audience

This publication is designed as a modular guide for criminal justice reform practitioners. Its approach grew out of the OSCE Regional Trial Monitoring Project¹, funded by the European Commission, where trial monitoring provides a structured empirical window onto institutional practice around processing serious and complex organized crime and corruption cases. The Guide builds on that experience by treating practice-based evidence as a starting point for policy work.

Accordingly, this Guide centres on what are often the most visible elements of the work of criminal justice institutions, namely active proceedings; specifically, through the work of courts and prosecutors' offices. While adaptable to a range of contexts, this process does not attempt to cover every institution, instrument or path through which rule-of-law or criminal justice reform can be designed.

As such, this is a justice-reform application of the policy-cycle logic, providing a disciplined method for making policy decisions in a field where legal standards, institutional behaviour, political sensitivity and imperfect evidence constantly interact. This Guide, therefore, speaks to two primary professional communities that often approach reform differently: legal practitioners on the one hand, trained to work with rules, standards, evidence and legal reasoning, and policy practitioners on the other, trained to work with uncertainty, trade-offs, feasibility, incentives, sequencing and implementation. Justice reform requires both disciplines. Legal certainty without policy realism produces reforms that may be technically correct but unusable. Policy pragmatism without legal discipline produces reforms that may be feasible but normatively thin. Recognizing though that comprehensive reform is usually the work several disciplines, it may also be useful to a wider range of actors: such as monitoring and evaluation practitioners, institutional advisers, data analysts, programme managers, civil society researchers, and others involved in diagnosing justice-system problems, designing reform responses, or assessing whether reform has turned into practice.

The Guide is intended to help practitioners understand enough of each other's logic to work from a shared method, when called to manage an issue faced by many involved in public sector-reform: how to turn evidence about what institutions actually do, into relevant reform that can be designed, implemented, tracked and revised; without jeopardizing core institutional functions in the process. The Guide answers this through a policy engine comprising a six-Step, closed-loop sequence that carries a body of evidence from practice observation through reform recommendation design to verification of its implementation and observable practice change.

¹ <https://www.osce.org/WesternBalkansTrialMonitoring>

II. Guidance on use

a) The six Steps

The engine operates through six main Steps, each moving logically into the next, with the sixth Step laying the foundation for returning to the first, creating a feedback loop:



The six Steps organise policy work around practice-based evidence on how cases are processed. They build on the broad functions familiar from trial monitoring — information gathering, analysis, advocacy and follow-up — while fleshing out the methodological components that become necessary once trial-monitoring evidence is used not only to identify and communicate reform needs but also to design, select and later test a policy response: that is, defining the evidentiary question, establishing what the evidence supports, testing why a pattern persists, designing a reform response and assessing what happens after implementation begins.

Although the Guide is presented in distinct Steps, it should not be read as six separate methods but as a single operating cycle. Each Step emerges from the previous one; as a reform recommendation's strength depends on the validity of the underlying diagnosis, the quality of the diagnosis hinges on the supporting evidence and analysis, and reform tracking is meaningful only if the expected practice change was clearly defined at the recommendation stage, and so on and so forth. For methodological reasons, it is advised to read the Steps in sequence, so practitioners can see how the process develops, even if they later return to use only one Step.

The policy engine metaphor has limits of course, as it does not produce automatic answers. We all must accept that reform decisions are rarely mechanical. They require professional judgement under conditions of incomplete evidence, competing values, institutional constraints and political sensitivity. The point of the engine is not to remove judgement, but to discipline it: to show what evidence was used, what inference was drawn, what uncertainty remains, what reform action is proposed, and what evidence would later show whether the reform is working.

b) Application conventions

For ease of use, and to give life to the methodology, the Guide applies several conventions.

First, it addresses a single reader, the “practitioner”, which stands in for anyone using the Guide, regardless of professional background or area of expertise. The Guide does not assume a single user; and indeed, it may often be deployed by a wider institution or organization. This term is merely a device for providing clear instructions.

The second convention appears in the margin notes that recur throughout the Guide. Each type of note serves a distinct purpose:

	Practitioner notes	Highlights decision points and judgement points to be made by the user.
	Method notes	Explains how to use a specific tool, sequence or analytical operation.
	Adaptation notes	Included at the end of each Step, elaborates on the use of each Step beyond organized crime and corruption cases, noting what must be re-specified when the method is applied to other case typologies, institutional sites or justice system functions.
	Ethical frame notes	Included at the end of each Step, identifies critical standards and risks implicated by the Step in question, drawing on principles established in the ODIHR Reference Manual and adding Guide specific safeguards.
	Communication cards	Included at the end of each Step, shows how communication is not an end-stage reporting task but a continuous interface with the reform environment, so each card sets out what to communicate at that point, and what the exchange should yield.
	RTM application notes	Illustrates a concept or tool and grounds the method by explaining its application during implementation of the RTM Project.

Third, the names of the six Steps — Scope, Collect, Analyse, Explain, Recommend and Track — are capitalized when referring to the defined stages of the policy engine. Lower case is used where the same words retain their ordinary grammatical meaning.

c) Organized crime and corruption as the worked application of the Guide

Because of the Guide's provenance in the RTM Project, serious and complex organized crime and corruption (OCC) cases are used throughout as illustrations. This does not limit the method to those cases. Rather, they provide a rigorous test of institutional practice as they concentrate pressures that can make strengths and weaknesses more visible. This includes complex evidence, multiple defendants, cross-border elements, financial questions, political or economic influence, media attention, and procedural contestation. In doing so, they provide evidence about whether the institutions and practices within the monitored scope maintain legality, fairness, discipline and consistency under demanding conditions. These cases also concern forms of criminality that OSCE participating States² have identified as threats to security and stability, giving their handling significance beyond the immediate professional or procedural context.

d) Adaptation to other fields and contexts

The tools on which this Guide draws were developed for use in the OSCE region, specifically South-Eastern Europe. Accordingly, the referenced legal mechanisms and procedural specifics draw on those legal frameworks, conventions and traditions. For example, references throughout the Guide to prosecutorial independence or autonomy are context-specific: the applicable domestic constitutional model governs, while international standards require safeguards against unjustified interference and the impartial exercise of prosecutorial functions. That does not, however, preclude the practitioner from applying the Guide in other contexts and to justice systems with other legal frameworks. This would require adjustments of course, as application in another context requires the practitioner to re-specify the applicable legal standards, institutional roles, observable decision points, source base and safeguards for that context. To support the practitioner, each Step provides guidance on how to adapt.

²MC.DEC/3/05; MC.DEC/5/06; MC.DEC/7/08 (Helsinki 2008); OSCE Strategy to Address Threats to Security and Stability in the Twenty-First Century, MC.DOC/1/03 (Maastricht 2003).

III. Sources and origins of this Guide's methodology

As noted above, this Guide draws on the OSCE RTM Project, which used trial monitoring to assess how serious and complex OCC cases are processed in South-Eastern Europe. The Project's trial-monitoring provided the structured evidence base for assessing how relevant criminal justice institutions function in practice in such cases. Within the wider approach developed through the Project, trial monitoring therefore serves as the principal evidentiary entry point rather than the end of the reform process. Building on that foundation, this Guide consolidates the Project's adapted analytical tools, Deep Dive methods, and practices for developing reform recommendations and tracking their implementation.

As such, this Guide does not reproduce a single methodology in full. Nor is it simply a descriptive account of the RTM Project. The approach draws on several methods, analytical disciplines and bodies of practice that shaped the RTM Project's implementation: traditional trial monitoring, socio-legal analysis, results-based management, policy design, institutional behaviour, policy implementation and impact evaluation, statistical analysis, multi-criteria decision reasoning, organizational learning and strategic communication. These disciplines are not presented as separate theories but are integrated into one coherent framework, because the practical problem at hand requires an integrated whole: evidence must be collected, interpreted, translated into reform, followed up, and communicated in a way that remains methodologically defensible, operationally usable, and institutionally credible. Different parts of the Guide make these methods operational but are not intended to replace specialized study of any of these fields. Practitioners who wish to go deeper should use the further-reading annex 2, which signposts core references for the main disciplines and methods used throughout the Guide.

Three principal sources inform this Guide:



OSCE RTM Project – Funded by the EC and implemented by six OSCE field operations in the South Eastern Europe, the RTM Project developed and applied detailed practical tools for trial monitoring, analysis, recommendation development and follow-up, and the practice is synthesized in this Guide.

ODIHR trial monitoring methodology, as adapted for the RTM Project: This internal methodology, developed by ODIHR in collaboration with the participating OSCE Field Operations and the OSCE Conflict Prevention Centre and tailored for the RTM Project, serves as the primary methodological basis for this Guide's Steps 1-3.

ODIHR's Trial Monitoring: A Reference Manual for Practitioners (2012) – The foundational OSCE reference guide on trial-monitoring principles and methods, and the baseline for the RTM Project-specific trial-monitoring methodology.

INTRODUCTION

Effective Criminal Justice Reform: An Evidence-Based Engine

I. Understanding criminal justice reform

For the purposes of this Guide, reform is considered to have taken hold when the targeted way of working or institutional conditions are observably different. New laws, strategies, guidelines, training programmes, co-ordination bodies and digital tools may all be important components of reform, but their adoption or delivery does not by itself demonstrate that the intended change has reached institutional routines. The practical test is whether institutions and professionals begin to perform their functions differently. On that basis, this Guide offers criminal justice reform practitioners a framework for using evidence about how justice institutions actually operate, to design, implement, track and assess related reform. It begins not with what a reform plan says should happen, but with what can be observed in the functioning of the system.

In criminal justice, much of that institutional activity is visible in proceedings and in formal legal acts issued by courts and prosecution authorities. Proceedings show how rules, authority, time, evidence and rights are managed; formal acts show how public authorities frame, reason about and record the exercise of their powers. Together, as the evidentiary basis of trial monitoring, these sources provide a disciplined entry point to the central question of this Guide: how can observable institutional performance inform analysis and explanation, the design of reform recommendations, implementation tracking and, ultimately, change and learning?

Grounded in this evidence, the Guide asks how law and institutional mandates operate in practice, what recurring implementation problems arise, and what those problems may indicate about institutional routines, incentives, capacity and reform needs. This discipline is particularly important in politically sensitive and high-visibility fields, where competing perceptions, institutional interests and contested narratives can shape both the diagnosis of a problem and the reforms proposed in response. A strong evidence-base helps guard against predetermined diagnoses, test contested assertions and create a more credible basis for institutional dialogue and public discourse. The aim is not to assess the outcome of individual proceedings, relitigate the merits of a case, conduct extensive doctrinal analysis, or mechanize reform decisions. Recognizing that public sector reform at large requires professional judgement under conditions of uncertainty, institutional sensitivity and incomplete evidence, the framework instead makes that judgement more disciplined and accountable: it shows what was observed, what was inferred, which pattern was identified, which contributing conditions the evidence supports, which reform response followed, and what evidence would later demonstrate whether reform implementation changed the targeted way of working.

II. Bridging the evidence-to-reform gap

In many reform settings, design, technical assistance, monitoring and evaluation are often, in practice, fragmented, and what is learned at one point in the reform cycle may not always inform subsequent stages. A monitored pattern may become a report finding without becoming an implementable reform. A recommended reform action may name a desired improvement without identifying the precise practice or decision point that must change. A reform activity may be delivered, while the evidence used to assess whether practice changed follows a different logic. The result can be a collection of assessments, action plans, training, follow-up meetings and public messaging that lack coherence and discipline.

The policy engine addresses this fragmentation by treating reform as a connected chain of reasoning rather than a sequence of adjacent activities. Each of its six Steps answers a distinct question and produces the basis for the next.

What matters, however, is not simply the sequencing of familiar operations. The Guide's distinctive contribution lies in the purpose for which these operations are brought together. Trial monitoring is itself a multifaceted justice-reform tool. As reflected in established OSCE practice, independent observers can contribute to transparency and confidence-building; systematic monitoring can identify strengths and weaknesses in the administration of justice; and its findings can support recommendations, reporting, advocacy, and follow-up. The familiar trial-monitoring cycle of information gathering, analysis, advocacy and follow-up is organized around those functions.

The policy engine takes the practice-based evidence generated through monitoring and subjects it to a more explicit policy-design and implementation logic. Where trial monitoring evidence is intended to support reform design, the practitioner must make explicit which conditions appear to sustain the observed practice; what specifically must change; which actors hold the authority or leverage to influence it; which alternative interventions might alter those conditions; which option is technically sound, politically supportable and operationally feasible; how the selected measures are expected to work together through a theory of change and results chain; and what evidence would later distinguish institutional activity from actual change in practice. Trial monitoring findings are therefore treated not simply as the end-point of analysis or an object of advocacy, but as integral part of various phases of a policy intervention, whose logic, dependencies, trade-offs and implementation pathway must be made legible through an evidence-base.

This does not make the policy engine an alternative form of trial monitoring. The same body of monitoring practice can serve more than one reform purpose, and the appropriate methodological architecture aligns with the decision it is intended to support. The Guide provides guidance for practitioners that does not establish a new OSCE/ODIHR trial monitoring standard, nor does it supersede existing standards. As such, within the OSCE, the ODIHR methodology remains the framework for operating a trial monitoring programme. The Guide's six-Step policy engine is intended to translate evidence generated through that framework into explicit policy diagnosis, intervention design, and implementation verification.

III. The policy engine: operating logic and safeguards

The Guide codifies a closed-loop approach developed through practice, bringing evidence generation, policy design and implementation verification into a single, recursive method. Each Step constrains what can legitimately follow; later evidence can reopen earlier decisions; and the limits, assumptions and ethical boundaries of the inquiry remain explicit throughout.

This structure also provides a shared, effectiveness-focused and evidence-informed basis for reform dialogue, particularly in politically sensitive environments where high-profile cases may generate competing narratives or defensive institutional responses. By making the analytical lens explicit and shifting discussion from broad claims about "capacity", "institutional capture" or "reform stagnation" to observable professional and institutional practices, the engine gives relevant reform actors a common vocabulary for evidence-based problem-solving. At the same time, combining evidence generation, reform design and verification within one framework creates its own risks, particularly confirmation bias.

The engine must therefore be understood not only as connected and recursive, but as methodologically bounded and subject to safeguards.

In noting what how this engine operates, it must be emphasized that this analysis does not address the legal correctness of the adjudicated outcome and should never be used, even informally, to relitigate a case; rather, each phase of the proceedings, including the verdict, is treated as observable evidence of institutional practice.

a. How the policy engine works

The policy engine is the disciplined sequence through which practice-based evidence informs reform recommendations, whose implementation can later be verified. By beginning **Scope**, the reform question shapes the selection criteria and the observation design. **Collect** then assembles evidence from multiple source streams, with trial monitoring complemented by official and administrative data, structured consultations, media and contextual material, and other inquiry tools provided by the Guide. Working across hearing, case, cross-case and bounded system levels, **Analyse** then asks what the evidence supports. **Explain** tests which conditions may contribute to an established pattern and why relevant practices may persist. **Recommend** translates the explanatory account into one or more reform recommendations. **Track** asks whether those recommendations were implemented and whether the expected changes are visible in practice.

The engine then restarts. This feedback loop, which uses evidence of practice to assess reform and to frame the next inquiry, is what makes the engine more than a sequence of separate tools.

b. A recursive rather than linear process

While the Guide presents the engine as a sequence, in practice the work is recursive. Depending on the reform and the available evidence, one cycle may suffice: define a case pool; collect evidence; analyse patterns; identify and test plausible contributing conditions; develop recommendations; and track implementation. In other situations, the first cycle produces a signal that standard evidence can identify but not resolve. For example, a low conviction rate may prompt a question, but it does not, by itself, establish a problem or explain any underlying condition. Further analysis may examine charging, evidentiary, legal, procedural or other factors. In such instances, the Guide provides the discipline to return, at any point, back to Scope, Collect and Analyse, adjusting the reform question based on what has been learned along the way.

c. Addressing potential risks and bias

Combining evidence generation, reform design and verification within the same exercise creates risks of confirmation bias. Those risks increase when the same people interpret the evidence, design recommendations and assess implementation. Certain safeguards are especially important. First, Scope and Recommend should define the claim limits and the evidence that would indicate implementation or practice change; Track should first assess implementation and practice change against those criteria; if the observed pathway diverges, it should treat that divergence as evidence for reassessing the diagnosis or recommendation through the appropriate earlier Step.

Where the same organisation also supports implementation of the recommended reform, implementation support and subsequent verification should, where feasible, be separated at team or decision-making level.

Practitioners assessing progress should not be solely responsible for delivering the measure being assessed. Where such separation is not possible, the dual role should be made explicit, assessment criteria fixed in advance, implementation evidence independently verifiable, and material assessments subject to cross-team or external review.

This ensures each Step remains distinct, informing rather than manipulating the others. Second, the method should preserve distinct claim levels, contrary evidence and alternative explanations, so systemic conclusions do not emerge from a single proceeding. This also creates accountability by showing why a given decision was taken.

IV. Creating an evidence base through effectiveness-based trial monitoring

Having set out how the policy engine operates, the next question is what kind of evidence can sustain it. In this Guide, that foundation is provided by effectiveness-based trial monitoring, which creates a structured empirical window onto institutional practice and can be complemented by other evidence sources as the inquiry develops.

a. Trial monitoring as the foundation

Trial monitoring provides that foundation because proceedings and formal legal acts make institutional practice observable. They show how rules are applied, powers exercised, rights protected, evidence handled and cases managed over time. Systematic observation and documentation convert those manifestations into a structured evidence base from which recurring patterns, strengths, weaknesses and further lines of inquiry can be identified.

In this Guide, trial monitoring therefore serves as the principal entry point, not the end of inquiry, for examining institutional practice—not because it can answer every reform question, but because it provides a consistent, source-grounded record on which further analysis can build. It can identify recurring patterns, the recorded reasons for particular events, and plausible questions about the conditions behind them. Using those findings for policy design requires a further methodological task though: testing which legal, institutional, organizational, professional or external conditions plausibly contribute to the pattern, and, where it persists, what helps reproduce it over time. A recorded immediate reason may explain why a particular event occurred without explaining why the same practice recurs across cases or survives reform efforts. The purpose is not ordinarily to establish strict causality, but to develop a sufficiently supported account of the conditions a reform response would need to address. Without that Step, a reform intervention may respond to the visible manifestation of a problem while leaving the conditions that sustain it largely unchanged.

b) Defining effectiveness

This Guide builds on traditional trial monitoring through the RTM Project's application of effectiveness-based trial monitoring (EBTM). It defines EBTM as a structured, mixed-methods socio-legal methodology that observes and analyses adjudicative processes and related institutional practice to assess, within defined limits, how effectively justice institutions fulfil their mandates and protect rights, and to generate evidence-based findings, explanatory accounts and reform recommendations. EBTM retains the core principles and methods of trial monitoring (including its focus on fair trial standards) while placing hearing observation and the review of formal judicial and prosecutorial acts within a broader inquiry of institutional effectiveness.

The pairing of fairness with effectiveness is long established in OSCE commitments. In the Brussels Declaration on Criminal Justice Systems,³ OSCE participating States identified competence and diligence as prerequisites to the due performance of judicial office, called on prosecutors to discharge their duties fairly, consistently and expeditiously, and concluded by calling for the fair and effective operation of criminal justice systems. The Guide's dimensions of Capacity, Fairness and Efficiency operationalize aspects of those commitments. It therefore examines not only whether defendants' rights are respected, but also whether serious and complex cases are handled with the capacity, fairness and efficiency required for the justice system to fulfil its functions effectively.

Effectiveness though is a broad concept, open to different interpretations. While this flexibility is useful, it is methodologically unworkable. Therefore, drawing on the RTM Project, this Guide uses a functional definition of effectiveness.

Definition — effectiveness

SOURCE



In very general terms effectiveness measures whether or not the institutions in question are achieving the goals society has set for them. In relation to this, one can distinguish between the general goal of the criminal justice system, which is to deliver justice by convicting and punishing the guilty while protecting the innocent, and the specific goals of the different institutions that constitute the system.

Source: ODIHR-developed TM Methodology, specifically tailored for the RTM Project, summarized.

The ODIHR formulation states the system's central task in its narrowest and most testable form. The purposes of criminal justice are of course broader than conviction and acquittal. Deterrence, rehabilitation, retribution, incapacitation and restoration have long been recognized as the ends a penal system is meant to serve, and they remain the reason societies invest in courts and prosecution services at all. This Guide takes those purposes seriously but does not attempt to measure them. Whether sentencing deters, whether punishment rehabilitates, whether communities are restored: these are population-level questions, shaped by economic conditions, policing, social policy and a great deal else that lies well outside the courtroom and beyond the reach of trial monitoring evidence. Nor does the Guide ask whether a particular verdict was correct. It examines the conditions under which that determination is made and leaves the determination itself untouched. What trial monitoring can establish is whether institutions perform the functions through which those broader purposes are pursued: whether proceedings reach a lawful and durable resolution rather than systematically altering on appeal, whether sentencing is reasoned and applied within the legal framework, whether proceeds of crime are traced, secured and recovered, and whether the procedural rights of all actors are observed.



RTM application

The RTM Project applies trial monitoring with a particular focus on the performance of public justice institutions and officials, notably prosecutors, judges, prosecutors' offices and courts. This reflects the Project's mandate and its broader connection to OSCE commitments concerning the functioning of public justice institutions. The RTM Project did not assess defence counsel performance as a separate institutional-performance domain. Defence conduct was considered only insofar as it was relevant to assessing the performance of authorities or to a fair-trial safeguard, for example equality of arms, adequate time and facilities, or the court's duty to ensure the fairness of proceedings. This decision pertained specifically to the RTM Project scope and does not represent a rule to be applied in every adaptation.

³ MC.DOC/4/06, Brussels, 5 December 2006.

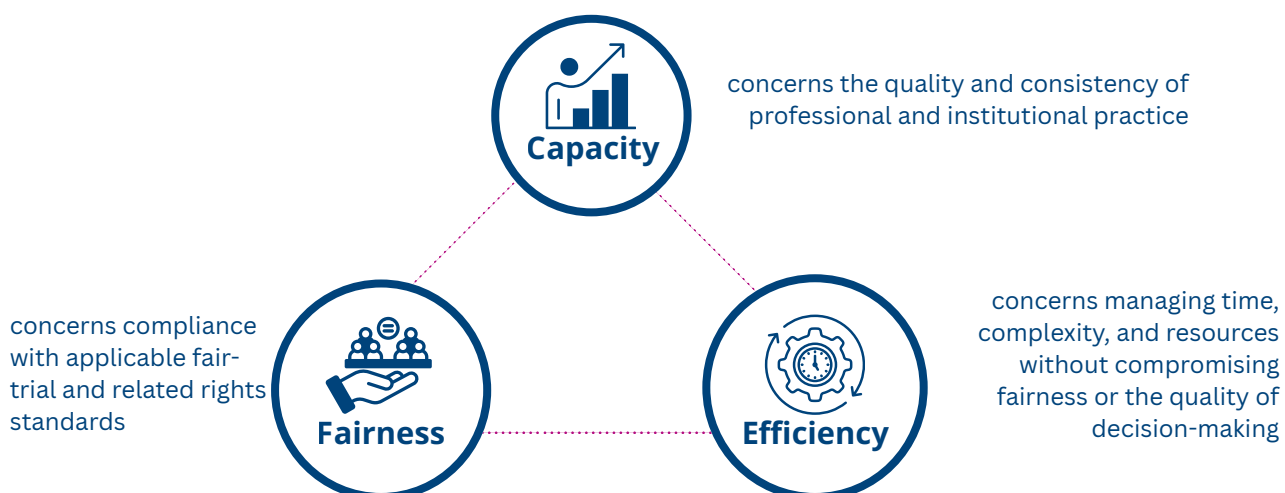
Effectiveness in these functions becomes evident through repeated working practices, which this Guide uses both to identify patterns and trends and to assess whether relevant practice changes following reform. A recurring pattern across comparable cases may raise questions about capacity, incentives, co-ordination or legal practice; those possible contributing conditions are tested in Explain. The practitioner therefore looks for repeated, observable manifestations of institutional performance while respecting the limits of the monitored pool. Recurrence across comparable cases may support only a bounded cross-case pattern; establishing a broader systemic claim requires substantial additional evidence on scope, distribution and institutional reach, and even then, it may be difficult or impossible to establish true causality.

A justice system may appear functional when cases are comparatively simple, stakes are low, and public attention is limited. High-pressure cases provide a demanding test of performance. Drawing on the RTM Project, the Guide treats serious and complex OCC cases as a purposive stress-test, asking whether relevant institutions maintain legality, fairness, consistency and procedural discipline under demanding conditions.

c) The dimensions of effectiveness: Capacity, Fairness and Efficiency

As discussed above, the RTM Project takes the effectiveness of the judicial response to serious and complex OCC cases as the primary object of EBTM inquiry. Because effectiveness is a broad and flexible concept, it must be translated into clear analytical dimensions. In fact, the range of potentially relevant issues is too broad to examine comprehensively, making deliberate selection and prioritization necessary. This is essential to provide a transparent basis for findings that may be subject to scrutiny, particularly in politically sensitive cases.

In defining those dimensions, the Project sought to balance the interest in effective criminal accountability with the protection of fair-trial rights, while also capturing both qualitative and quantitative aspects of judicial performance. On that basis, EBTM examines effectiveness through three interrelated dimensions:



Fairness is constitutive of effectiveness in the framework of this Guide: gains in speed, output or enforcement do not amount to effective performance where applicable fair-trial guarantees or judicial independence are compromised. The content of each dimension is presented under Step 3-Analyse. Based on EBTM and its defined dimensions, the Guide systematically translates collected evidence into policy-relevant findings and explanatory questions; not merely as a monitoring technique but a diagnostic method that provides bounded empirical feedback on justice-system practice.

d) Why evidence can create reform momentum

Nothing in the sequence from Scope to Track itself drives change or compels an institution to act, but the policy engine can improve the conditions for reform dialogue by combining evidence with mechanisms intended to support shared problem definition, follow-through, transparency and learning.



The diagnostic mirror

Practice-based evidence can reveal a gap between formal commitments and observed performance, shifting the discussion from broad claims about laws, capacity or political will towards specific institutional practices. Its value lies in making those practices sufficiently visible and bounded to enable discussion.



Shared problem-solving

Once a bounded pattern is visible, dialogue can shift from general attribution to more precise questions: which conditions may contribute to the pattern; which decision points they operate at; which actors hold relevant authority; and which entry points should be assessed. A shared diagnostic vocabulary can provide a more precise basis for reform dialogue in politically sensitive fields.



Follow-through

Structured dialogue and periodic review can help translate findings into recommendations and recommendations into implementation decisions, with clear responsibility, sequencing and evidence of uptake. Without follow-through, a finding does not become a tested reform proposition or implemented response.



Sequenced transparency

Where the mandate, confidentiality obligations, security risks and the independence of the monitoring exercise permit, technical engagement, including stakeholder consultations, can help institutions understand the evidence, identify factual errors and consider possible responses. Such engagement must not give an institution a veto over findings. Subsequent public reporting may then strengthen transparency and accountability.



Learning and adaptation

A reform that works in one setting may inform practice elsewhere; a recommendation that cannot be verified may be narrowed or reformulated; a persistent pattern may trigger a Deep Dive, a tool further elaborated in the Guide in Annex 1; and a pattern that no longer appears may be deprioritized after sustained verification. Track does more than assess implementation: it generates the next set of questions and turns evidence into a cycle of learning.

Step 1



► Where you are in the cycle

The policy engine turns evidence on institutional practice into reform that can be designed, implemented, assessed and revised. Scope is the first operational Step of this process, as it determines which practices need to be examined, which cases or proceedings will be included in the evidence-base, how closely they will be monitored, and which claims the resulting evidence may support.

► Why this Step is important

Sound, relevant and actionable recommendations require an evidence-base tailored to the reform question. Scope applies a purposive case-selection strategy: cases are selected for their diagnostic value, not to estimate prevalence across the full caseload. Scope produces an evidence strategy that defines the eligible field, priority criteria, observation window, monitoring modality and depth, expected sources and inference limits, allowing later findings to be traced back to the design of the evidence base.

► What happens if we are not intentional about it

Without deliberate scoping, the evidence-base becomes unstable. A scope that is too broad produces shallow coverage; one that is too narrow may yield isolated or highly conditional observations. A pool driven mainly by access or visibility can replicate the justice system's own blind spots.

► What you produce

By the end of this Step, the practitioner should have:

the reform question; a context baseline; subject-matter, institutional, geographic and temporal boundaries; eligibility and priority rules; an inclusion/exclusion register; case-level selection notes and monitoring-modality assignments; observation windows; claim-to-source and access maps; evidence-gap and risk records; an approval record; a version history; and a review schedule.

THE STEP AT A GLANCE

1

Articulate the reform question

Why are we doing this at all?

Turn the reform concern into a neutral, answerable question about institutional practice: broad enough to justify sustained monitoring, precise enough to guide data collection, and agnostic to the finding you already suspect

2

Map the context baseline

What do we already know?

Prepare a short context baseline: legal framework, institutional structure, available statistics, known reform gaps, stakeholder map and likely evidence gaps.

3

Define the eligible field

What qualifies for monitoring, and why?

Define the substantive field of observation by setting the subject matter, institutional site, jurisdiction, geographic and temporal coverage and the units of observation and analysis.

4

Identify the monitored case pool

Which eligible cases matter most?

Define the criteria for selecting the cases to be monitored based on where institutional practice becomes diagnostically visible and valuable.

5

Set the observation window

When do we start and stop monitoring a given case?

Decide at what procedural stage a prioritized case can enter the evidence base and how far it should be followed.

6

Calibrate monitoring modality and depth

How should the case be monitored?

Choose full or Tier 2 monitoring and set the coverage and source intensity needed for the intended claim.

7

Write down the evidence strategy

What must be collected, why and how?

The evidence strategy tells the evidence collection process what to observe, request, structure for analysis and review

► What Scope does: from reform concern to evidence strategy

Scope converts a reform concern into an auditable evidence strategy. It provides a plan for conducting a rigorous inquiry into the reform question through systematic monitoring. In doing so, it defines the field of observation, the cases or proceedings that may enter the evidence base, the basis for prioritization, the observation window, the monitoring modality and depth, and the limits of permissible inference.

A monitoring exercise usually begins when an aspect of justice-system practice warrants examination for possible reform. Serious corruption cases may appear to take too long; confiscation may be rare; case outcomes may appear to vary by defendant status; or plea agreements may resolve complex cases with sentencing patterns or reasoning that warrant closer examination. At Scope, these are concerns to investigate, not conclusions to confirm.

Scope therefore poses a neutral reform question, records the context baseline, defines the eligible field, defines criteria to prioritize cases for diagnostic value, sets the observation design, and documents the sources and limits that Collect must carry forward. Together, these decisions constitute the completed evidence strategy.

Scope does not predetermine findings. A purposive pool may show that a pattern recurs in the monitored material; it does not, by itself, establish prevalence across the system's full caseload. A later systemic claim must be confined to a defined institutional or procedural scope and supported by evidence of institutional reach - for example, a common rule, workflow, routine or arrangement demonstrated through appropriately claim-fit caseload, documentary, multi-site or other evidence.

1 Articulate the reform question

Scope begins with a reform question: a neutral, answerable question about institutional practice that explains why the evidence base is being built and what claims it must support.

The question should be broad enough to justify sustained monitoring and precise enough to guide selection. A question that is too broad makes almost every case relevant. A question that is too narrow does not allow for meaningful understanding, as it may lock the exercise onto a single occurrence, institution or procedural feature before the evidence shows whether the issue recurs.

The practitioner must be careful when framing the question to avoid embedding suspected findings from the outset. Questions such as "Are courts biased against powerful defendants?" or "Does confiscation fail because judges lack knowledge?" invite confirmation bias and premature attribution. Frame the observable practice or knowledge gap first; analysis and explanation will determine what the evidence supports.

A sound reform question defines the field of inquiry without asserting an answer. It should allow the evidence to identify concerns, variation, good practice, or the absence of a material problem.

A **useful reform question** should do four things:

First, identify the practice area and the institutional site. In RTM, this is the adjudication of serious organised crime and corruption proceedings and the performance of courts and prosecutors' offices.

Second, identify the performance concern. Applying the EBTM framework will help identify concerns related to fairness, efficiency, capacity, or how they interact.

Third, identify the applicable legal or policy benchmark, where one exists, and state separately why the issue matters for reform. Equality before the law, the right to a trial within a reasonable time or legal certainty may supply legal standards in the relevant context; public confidence, institutional resilience or implementation of formal reforms may instead be reform rationales or objectives and require evidence appropriate to those claims.

Fourth, remain open to testing and state the intended scope of the claim and recognise the limits of the scope. The question should define what must be observed without assuming a defect, a cause, or a responsible actor. A single occurrence cannot answer a question framed as a systemic inquiry.



RTM Application

The RTM Project expressed the central reform question as follows:

How effectively does the justice system handle serious organized crime and corruption cases from indictment through final resolution, in terms of capacity, fairness and efficiency, particularly where proceedings involve high-status defendants, substantial alleged harm or significant evidentiary and procedural complexity?

This question offers an example of how to identify the field and performance pressures without assuming dysfunction or success.



Practitioner Note

Scope should begin with one primary reform question stated in a single sentence; narrower sub-questions may be added where they are necessary to operationalise it. Later decisions about boundaries, prioritisation, observation and sources must all answer the same question: what must the evidence base be capable of showing?

2 Map the context baseline

Once the reform question has been named, the practitioner maps the context baseline.

The purpose is simple: before deciding which cases to monitor, the practitioner must understand the policy and practice landscape into which monitoring will enter. Scope should not begin in a vacuum, but with a disciplined account of what is already known, what is assumed, what is contested, and what the monitoring exercise still needs to test.

The context baseline records the applicable framework, institutions, available data, previous reform work, access conditions, relevant stakeholders and initial evidence gaps. It serves as a scoping input, not a set of findings, and is not the tracking baseline used later for Track.

For each source, record its producer, date, provenance, definitions, coverage and known limitations, including any incentives that may affect self-reported data. Mark information as verified, contested, incomplete or unverified. Do not carry forward conclusions from earlier assessments without verifying their evidentiary basis and continued relevance.

A context baseline should cover at least the following:

01 Baseline element:

Legal and procedural framework

What the practitioner records:

Applicable criminal-law, procedural and human-rights standards; specialized tools; material legal changes; and the source and date of each rule.

02 Baseline element:

Institutional and jurisdictional map

What the practitioner records:

Institutions, courts, prosecution offices, appellate or enforcement bodies, their mandates and the geographic or jurisdictional boundaries relevant to the inquiry.

03 Baseline element:

Available evidence and statistics

What the practitioner records:

Producer, date, unit, definitions, coverage, denominators, disaggregation, known incentives and limitations. Mark the information as verified, contested, incomplete or unverified.

04 Baseline element:

Known reform work

What the practitioner records:

Previous assessments, strategies, recommendations and evaluations, including their evidence base, date, status and continued relevance.

05 Baseline element:

Case pool and flow

What the practitioner records:

The known or estimated eligible field, case inflow, procedural distribution and any limits on identifying the full eligibility field.

06 Baseline element:

Access conditions

What the practitioner records:

Likely access to hearings, legal acts, registers, administrative data and institutional counterparts, including restrictions and contingency routes.

07 *Baseline element:***Stakeholders and decision authority**

What the practitioner records:

Actors who hold relevant knowledge, control access, face risk, or have authority to act on eventual recommendations.

08 *Baseline element:***Public and media context**

What the practitioner records:

Public narratives, official statements, transparency issues and media attention that may affect visibility, safety or public salience.

09 *Baseline element:***Initial evidence and risk gaps**

What the practitioner records:

What is not known, cannot be verified, is only indirectly inferable, may be obtainable from another source, or is unavailable; and any associated ethics or safety risk.

3 Define the eligible field

3.1 Define the subject-matter boundary

Once the reform question has been named and the baseline context mapped, the practitioner defines the subject-matter boundary and the eligible field before choosing individual cases. The field should specify the subject matter, institutional site, jurisdiction, geographic and temporal coverage, target eligible caseload, where applicable, unit of selection, unit of observation, and intended unit or units of analysis. These boundaries determine the reach of every later claim.



RTM Application

For the RTM Project, the eligible field comprised organized crime and corruption proceedings within the relevant courts and jurisdictions. This is one application of the general scoping method, not a universal boundary.

The subject matter delineation

SOURCE



The pool is organized crime and corruption (OCC), identified primarily by domestic definitions in the relevant criminal codes: the core corruption offences of active and passive bribery, and corruption-related offences such as abuse of office, trading in influence, embezzlement and money laundering. Offence lists are harmonized across jurisdictions so far as possible, to allow comparable data. For organized crime, only cases in which the accused is formally charged with a recognized form of organized crime – criminal association, conspiracy, participation in an organized criminal group – fall within scope, regardless of the underlying offence. Where a domestic offence is doubtful, reference is made to international standards. The ODIHR methodology (2012) also required offence lists to be harmonized across jurisdictions so far as possible, with reference to international standards where the domestic classification was unclear.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

3.2 Write the eligibility rule and resolve borderline cases

01

Boundary Element:

Included offences or proceedings

What must be decided:

The domestic offences, proceeding types or decision categories within scope.

02

Boundary Element:

Organised-crime or corruption related offence rule

What must be decided:

Whether a formal organised crime or corruption-related charge is required and when related offences such as money laundering or embezzlement qualify.

03

Boundary Element:

Excluded matters

What must be decided:

What falls outside the field even if serious, visible or connected to an included case.

04

Boundary Element:

Institutional site and jurisdiction

What must be decided:

Which courts, prosecution offices or other adjudicative bodies are within scope, including competence and mandate limits.

05

Boundary Element:

Geographic and temporal coverage

What must be decided:

The locations and project period covered, including treatment of cases begun before or continuing after that period.

06

Boundary Element:

Target eligible caseload

What must be decided:

The known or estimated eligible caseload and any limits on identifying it.

07

Boundary Element :

Units

What must be decided:

The unit of selection, unit of observation and intended unit or units of analysis.

08

Boundary Element:

Borderline classification

What must be decided:

The reference hierarchy, legal reviewer, approver and record used to resolve doubtful classifications.

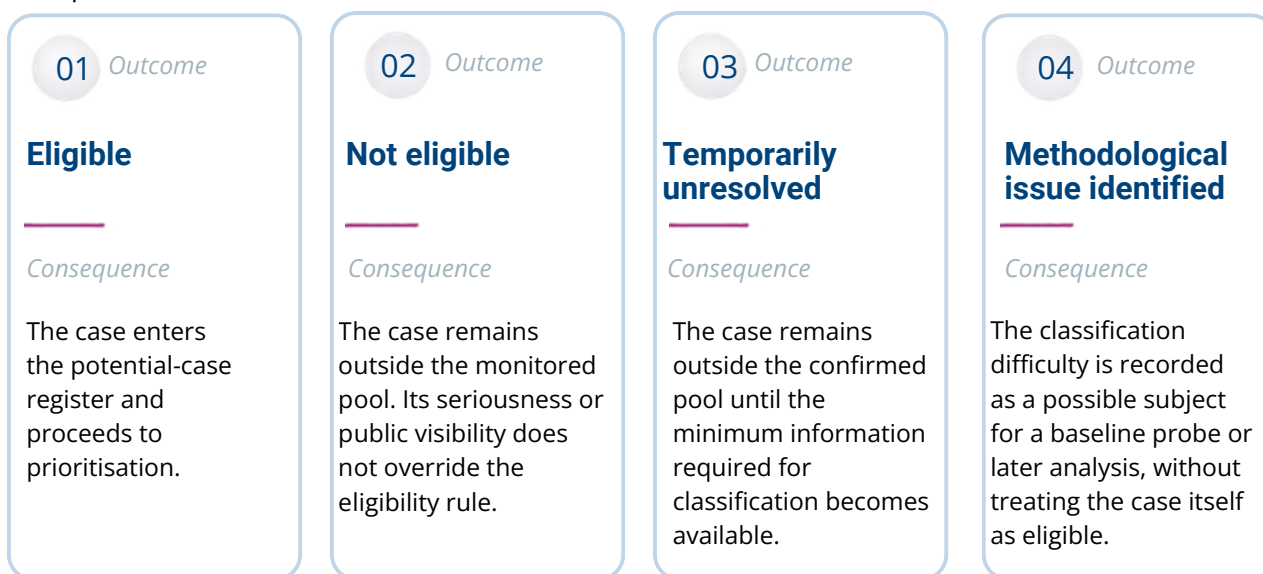
Translate the determined field into an eligibility rule that can be applied consistently. Some cases may substantively concern organized crime or corruption even where the formal charges do not fall within those categories. Whether such cases should be included in the monitored pool depends on the project's mandate and objectives. The RTM Project adopts a formal approach, generally limiting the pool to cases in which organized crime or corruption is reflected in the legal classification of the charges. Other monitoring projects may reasonably adopt a more flexible approach, allowing inclusion where the underlying conduct clearly engages organized crime or corruption concerns, provided the criteria are defined in advance and applied consistently.

In general, borderline cases must be decided through a regularized approach rather than on intuition.

This begins by recording the following:

- the domestic charge;
- the conduct described in the indictment or other available act;
- the possible functional connection to the subject matter;
- the agreed domestic, comparative or international reference point;
- the reason for inclusion or exclusion.

The possible outcomes are:



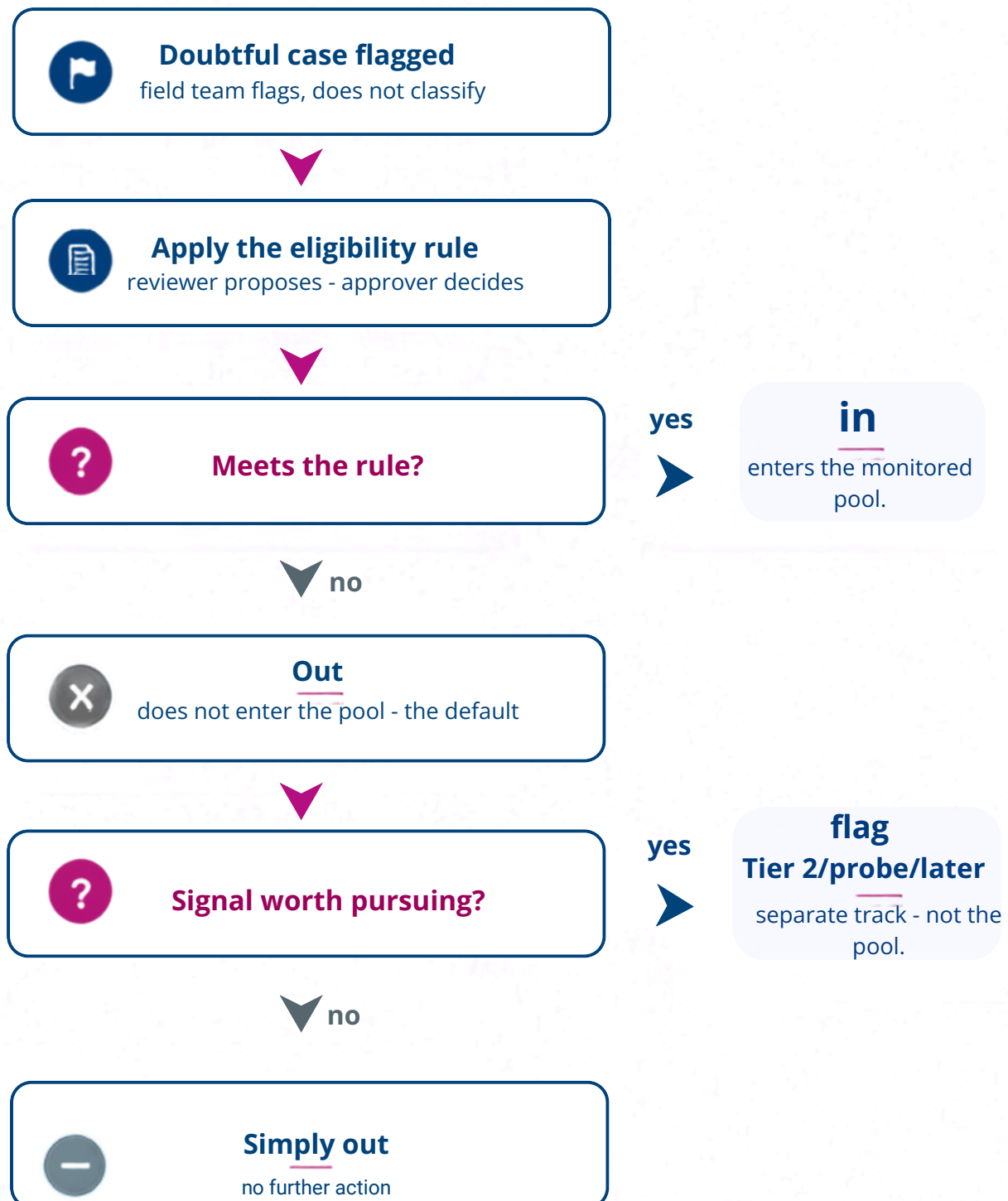
Where the methodology requires a formal charge, apparent substantive similarity is not enough to establish eligibility. A case concerning organized conduct but lacking the required organized-crime charge may therefore be excluded from the monitored pool. Its exclusion may nevertheless reveal a separate question about charging or classification practice. That question may justify a baseline probe, a Deep Dive or formal re-scoping. It should not enter Tier 2 unless the approved field is first amended through Scope and the methodological basis for that change is recorded.



Method Note

The monitored pool may support bounded findings about recurring practice within its stated scope. Population-level or whole-system claims require additional caseload, documentary, or multi-site evidence.

In a snapshot this would be the decision path:



Every decision is logged to the preceding log — checked before the next borderline case, so like cases are decided alike.

4 Identify the monitored case pool

Once the eligible field is defined, determine which eligible cases offer the greatest diagnostic value for the reform question. This is purposive prioritization, not probability sampling. The aim is a pool that is focused enough to reveal relevant practice, broad enough to test recurrence and variation within its stated scope, and transparent enough to withstand scrutiny.

4.1 Identify the diagnostic value

A case or group of cases has diagnostic value when it makes the practice relevant to the reform question sufficiently visible for the intended analysis.

The following **factors** guide professional judgement:

- **Reform question** – What institutional performance problem is being tested;
- **Formal mandate** – Which legal or policy area the project is authorised to examine;
- **Observable practice areas** – Where practice can be captured;
- **Comparability need** – Whether similar cases or decisions can be identified across courts, regions, institutions or jurisdictions;
- **Feasibility constraint** – Whether the project can access enough material to sustain analysis;
- **Public interest or rights implications** – Whether the practice bears on fundamental rights, vulnerable parties, or public confidence in institutions;
- **Repeated discretionary inflection points** – Whether the case type turns repeatedly on choices the rule does not fully determine (charging, detention, admissibility, adjournment, sentencing, recovery);
- **Use of specialised tools or procedures relevant to the type of case** – Whether the case involves the use of specific mechanisms or tools such as financial investigation for economic crime cases.

This prevents the case pool from becoming a collection of interesting files. It also protects the exercise from possible integrity doubts that the sample was cherry-picked, politically motivated or analytically incoherent.



Practitioner Note

A finding is only as defensible as the pool from which it is drawn. If the selection logic is unclear, every later conclusion becomes easier to contest. Scope therefore needs to be written down before monitoring begins, revised transparently when needed, and applied consistently enough that another practitioner can understand why a case was included, excluded or monitored at a different depth.



RTM Application

The RTM Project prioritised serious OCC cases through status of the defendant, gravity of the offence and media attention, while recording complexity separately. These attributes serve different purposes: status concerns documented power; gravity concerns alleged harm and public consequence; media attention concerns public visibility; and complexity concerns the evidentiary and procedural demands of the case.

Complex and serious cases carry specific diagnostic value. This is not simply because they garner attention but rather because they demonstrate how the system performs under stress, rather than how it processes routine cases.

Such proceedings often combine several stressors:

- complex or voluminous evidence;
- multiple defendants or linked proceedings;
- financial gain and asset-recovery questions;
- specialised legal and procedural instruments;
- vulnerable victims or witnesses;
- coordination across investigative, prosecutorial and judicial institutions;
- contested legal or evidentiary issues;
- media and public attention;
- actors with public, political, economic or social influence.

That is why this pool is diagnostically useful. Simple cases may show whether a system can process routine work. Serious and complex cases show whether the system can maintain legality, fairness, reasoning, co-ordination and time discipline when the stakes rise.

4.2 Apply the prioritisation and scoring model

Translate the prioritization rationale into criteria that practitioners can consistently apply to individual cases. This should be done in two stages. The practitioner should first identify the substantive characteristics that make an eligible case particularly relevant to the reform question, as outlined above. The practitioner should then combine those characteristics using a weighted scoring model to determine monitoring priority.

The method is a simplified multi-criterion weighted ranking model: no single characteristic determines priority; several criteria are assessed separately, assigned explicit weights and then considered together. This makes the prioritization decision transparent, repeatable and open to review.

a) The substantive criteria

The ODIHR methodology translates the seriousness of an eligible case into four substantive criteria: the status of the defendant, the gravity of the alleged conduct and the level of media attention.

Status of the accused — relates to the level of power enjoyed by the accused because of his/her de jure or de facto position in the society, whether in public or private sector. Where a case has several defendants, the criterion is assessed on the highest-ranking accused.

SOURCE



Gravity of the conduct — relates to the gravity of the direct consequences of the offence on the victims, including economic damage. It takes the assumption that some sectors are particularly sensitive and/or prone to corruption (such as healthcare, public procurement, spatial planning, infrastructure, and education) and thus that cases of corruption in these sectors have an additional impact in terms of gravity of the conduct.

Likewise, this criterion specifically takes into consideration (and thus scores higher) the gravity of the damage caused to the specific victims of underlying criminal offence of the organised crime consisting of trafficking/smuggling in human beings, drugs, or firearms; money laundering; embezzlement; or financial crimes.

SOURCE



Significant media attention — relates to the level of coverage the case receives, in mainstream outlets and in smaller specialised ones alike. This criterion weighs less than the other two, and a case strong on status and gravity is never excluded for lacking it.

ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

The three criteria capture different forms of diagnostic value. Status tests whether institutional practice remains consistent where power or influence is present. Gravity captures the human, economic and institutional consequences of the alleged conduct. Media attention captures the public environment in which public scrutiny, transparency and institutional communication become particularly salient; it is not itself evidence of public confidence or population-level perception.

In organized crime and corruption proceedings, seriousness commonly brings complexity with it. High-status defendants, grave conduct and significant public interest are often accompanied by multiple defendants and charges, complex evidentiary chains, financial or digital evidence, cross-border co-operation, corporate and ownership structures, specialized legal tools and demanding co-ordination between institutions.

Complexity is therefore not treated as a separate scored criterion. It is an inherent feature of many of the serious cases selected through the methodology and one of the reasons they are diagnostically valuable. These cases test whether the justice system can manage not only high stakes, but also the procedural, evidentiary and organizational demands that those stakes ordinarily generate.

At Scope, these criteria do not constitute findings about the case. They explain why the case may be especially worth observing. Status, gravity and media attention may later reappear in Analyse as variables for disaggregation or cross-tabulation, but at this stage they remain descriptors of the case pool.

b) The scoring system

These criteria are then converted into a scoring model to enable consistent prioritization. This weighted ranking instrument helps practitioners distinguish high-, medium-, and low-priority cases and allocate limited monitoring resources in line with the reform purpose. High-priority cases are ordinarily prioritized. Medium-priority cases remain eligible and may be included where needed to achieve adequate coverage, variation or comparability, including when the available high-priority pool is too small or too concentrated. Low-priority cases are ordinarily excluded unless documented evidence of need or unusual diagnostic value justifies an exception.

Further, the criteria should not be amended simply because early outcomes are inconvenient or a stakeholder seeks a preferred pool; this is the very rigour the criteria should provide. Changes to eligibility, priority criteria, weights, thresholds or monitoring-modality rules must be made carefully and formally, based on evidence demonstrating the need for change.

SOURCE



Scores are assigned to each criterion to determine whether a case is overall high-, medium- or low-level. Namely:

Status of accused:

High+ (High ranking or profile elected public office holders): 4 points
High (other categories than high ranking profile elected public office holders): 3 points
Medium: 2
Low: 1

Gravity of the offence:

High: 3 points
Medium: 2
Low: 1

Significant Media Coverage

Present: 1 point
Absent: 0

Based on the sum of scores (max. 8, min. 2), cases can be categorized as follows:

High level: 6-8 points

Medium level: 4-5 points

Low level: 2-3 points

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project.

High-priority cases are ordinarily prioritized for monitoring. Medium-priority cases remain available to broaden or balance the pool where needed, including when too few high-priority cases exist to support the intended comparison or coverage. Low-priority cases are ordinarily excluded. An exception may be justified where a lower-priority case has unusual diagnostic significance under one of the criteria or responds to a clearly documented evidence need. Any departure from the ordinary prioritization rule should be recorded in writing.


Practitioner Note

The scoring model makes a set of methodological value judgements explicit. To avoid conflating the aggregate monitoring-priority band with the separate defendant-status category, report the aggregate score together with its component profile (for example, status 4 / gravity 1 / media 1).

First, status carries a maximum of four points, gravity three and media attention one. Status therefore carries slightly greater weight than gravity, while media attention plays a supplementary role. This weighting reflects the Project's reform question: whether legal and institutional routines remain effective, fair and consistent where the accused holds substantial public, political, economic or social power. A different monitoring exercise, designed around a different reform question, may legitimately choose different criteria or assign different weights.

Second, the model allows limited compensation across criteria. A particularly high score on status may offset a lower score on gravity, and vice versa. This is intentional: the aggregate represents monitoring priority, not a claim that the criteria are substantively interchangeable. Where a case has exceptional diagnostic value that is not adequately captured by the aggregate, the practitioner should document the reason for overriding the ordinary ranking. Where plausible alternative weights or thresholds would materially change a case's priority band, record that sensitivity and treat the score as a ranking aid rather than a precise measure of diagnostic value.

Third, the categories may travel forward as analytical variables, but their meaning changes. Status and gravity may later be used to test whether findings distribute differently across types of defendants or conduct. These variables may later be examined in Analyse or, where warranted, a Deep Dive to test whether status, gravity or media attention are associated with differences in case trajectories or outcomes. At Scope, however, the categories remain selection descriptors.

5 Set the observation window

After prioritizing the pool, define when each case enters observation and how long it will be followed.

The observation window is the portion of the proceeding from which the evidence base will be constructed. It is a temporal and procedural boundary, distinct from monitoring depth, answering:

At what procedural point does observation begin, what event ends it, and which later enforcement or implementation stages remain relevant to the reform question?

Criminal proceedings unfold through a series of closely interconnected stages, with developments at one stage shaping what happens at those that follow. A hearing may reveal a particular event, but its significance can only be understood considering what preceded it and the consequences it may have for the later course of the case. An indictment sets out the prosecution's legal and factual theory and frames the issues to be tested at trial. The trial record, in turn, provides the procedural and evidentiary context reflected in first-instance reasoning and outcome, while appellate proceedings may confirm, correct or reverse earlier findings. Assessing the proceedings therefore requires attention to the case as a whole, rather than to each stage in isolation.

A wider and more continuous window improves the practitioner's ability to trace the case's pathway. As larger monitoring windows also require more access, time and resources, Scope should define the intended window in advance, state the minimum expected coverage, and record all gaps or departures. This also ensures the same window is applied, wherever possible, to all cases, enabling consistency.

SOURCE



Under the ODIHR-developed methodology for the RTM Project, cases should, as a rule, be monitored from the indictment phase through to their conclusion.

Monitoring does not begin before the filing of the indictment. This boundary reflects the nature and safeguards of trial monitoring.

Where possible, the observation window extends from the indictment to the final and binding decision. Within that window, monitoring may include:

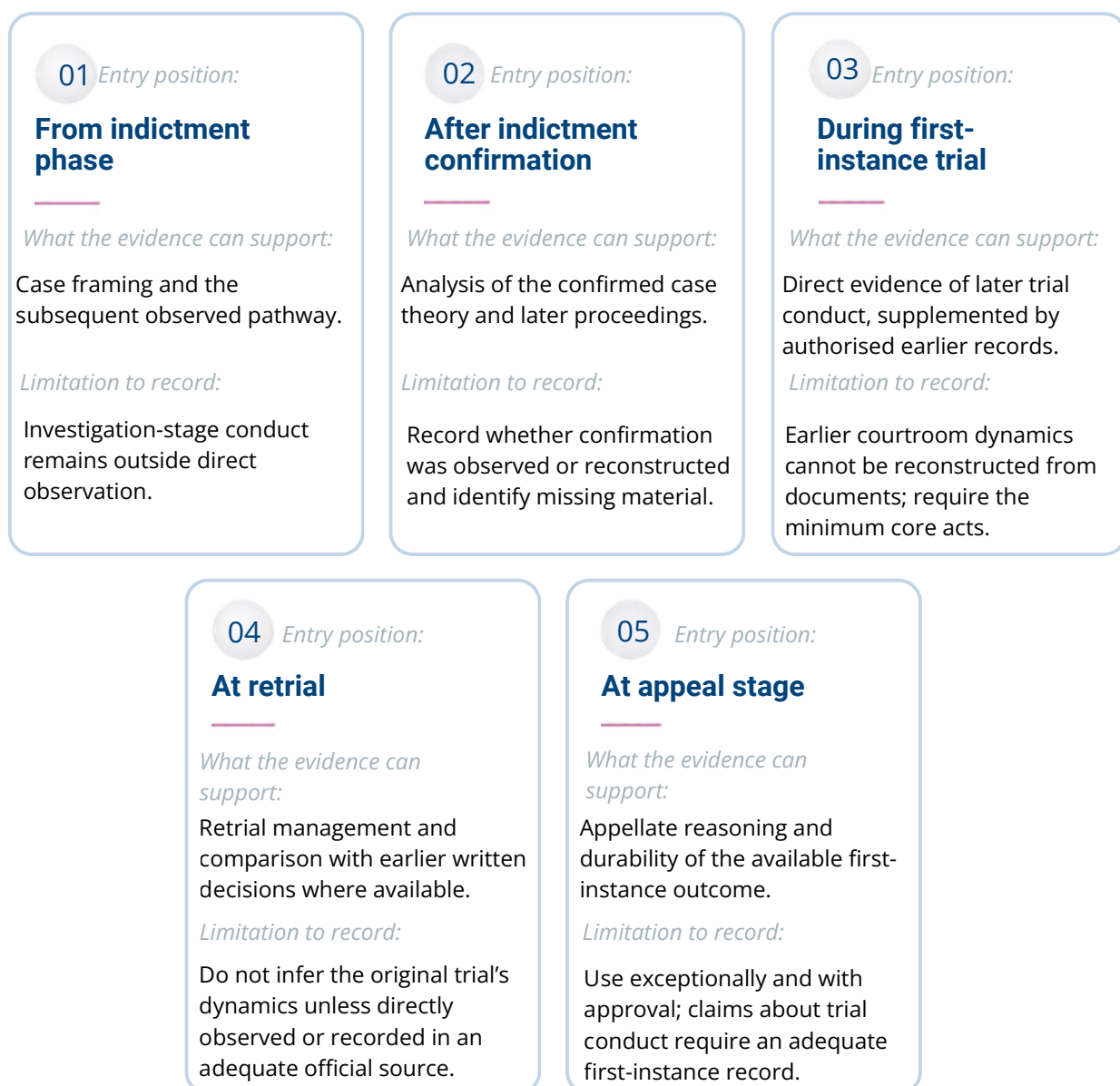
- review of the indictment and decisions concerning its confirmation;
- plea hearings and plea-agreement proceedings;
- first-instance trial hearings;
- procedural and evidentiary decisions;
- first-instance judgment and sentencing;
- retrial proceedings;
- appellate hearings and decisions;
- confiscation-related decisions and subsequent developments, where relevant and accessible.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project.

5.1 Entry into the observation window

The preferred entry point is the confirmation of the indictment (or, where domestic procedure does not provide for confirmation of the indictment, the filing, issuance and judicial review stage). This allows the practitioner to record how the prosecution formally frames the alleged conduct, legal qualification and theory of liability, and, where the applicable domestic framework requires it, the evidentiary basis relied upon, before observing how the court tests the case. The design deliberately excludes direct monitoring of the investigation stage; conclusions about pre-indictment conduct therefore require separate authorized sources. In addition, the next Steps in this analytical and reform exercise should account for, and be clear about, the lack of data before the indictment.

A case does not always become available at its ideal entry point, however. Monitoring may begin after the indictment stage or after the trial has already started. The matrix below explains the different possibilities:



Documents may bridge an entry-phase related information gap, including indictments, hearing records, procedural decisions, official transcripts, and other authorized material. This is documentary reconstruction, and should be presented as such, not a retrospective observation of hearings.

5.2 Following the case to its conclusion

Where feasible and within the approved observation window, a case should ordinarily be followed through to the final and binding decision, as many relevant data points emerge after the first-instance verdict. That said, monitoring itself commences at the indictment phase and does not wait until the final and binding verdict, both because relevant evidence can be extracted from each phase of proceedings and, given the length of proceedings, reliance on the final and binding verdict will create inevitable delays. More fundamentally, RTM does not analyse a first-instance judgement to pronounce on whether its outcome was ultimately correct – that is the function of appellate review. It analyses the judgement as an observable product of institutional performance: whether the reasoning identifies the issues, connects the facts, evidence and law, addresses material submissions, and is internally coherent. An appellate reversal may show healthy self-correction without erasing any independently supported upstream capacity concern identified in the first-instance reasoning.

Following a case beyond first instance may show whether:

- factual determination and legal qualification remain stable;
- reasoning is upheld or criticised;
- procedural defects lead to reversal or retrial;
- sanctions change on appeal;
- confiscation decisions are confirmed, altered or omitted;
- the result clarifies earlier concerns or good practices.

This continuity is particularly important for the End-of-Case Assessment (Step 3-Analyse). A final and binding judgement may conclude the monitored adjudicative pathway while leaving enforcement, confiscation or implementation questions unresolved. Given that, the observation window should not be treated as mechanically closed. Whether those developments remain within scope depends on the reform question and should be stated in the evidence strategy.

5.3 Changes in the procedural trajectory

Anticipate procedural changes that may alter the observation window, including:

- transfer of the case to another court or prosecutor's office;
- change in jurisdiction;
- severance or joinder of proceedings;
- change in panel composition;
- discontinuance or withdrawal of charges;
- plea agreements involving some defendants while others proceed to trial;
- retrial following appeal;
- loss of access to hearings or legal acts.

These events do not automatically remove a case from monitoring, but they may change the unit of observation or analysis. The practitioner should decide whether the original case remains the appropriate unit, whether linked proceedings should be tracked separately, and whether the change creates an evidence gap that must be recorded.



Practitioner Note

Continuity is an analytical asset. The purpose of the observation window is not to maximize the duration of monitoring for its own sake. It is to preserve enough of the case trajectory for later analysis to distinguish a procedural event from a recurring case-level trajectory.

6 Calibrate monitoring modality and depth

After fixing the observation window, choose the monitoring modality and set the depth needed to answer the analytical question. Modality determines whether the case is followed through full monitoring or Tier 2 document-based monitoring. Depth describes the coverage and source intensity required within the chosen modality to support the intended claim. Both are evidentiary choices and should be guided by the claim, required sources, procedural posture, diagnostic value, feasibility and risk, not by political visibility or the monitoring-priority score alone.

Within the chosen modality, set the minimum depth needed for the intended claim: identify which hearings or legal acts must be captured, the continuity required across the observation window, the source intensity needed to test the question, and any gaps that would make the claim unavailable or require it to be narrowed.

6.1 Matching modality to the analytical question

Not every case needs to be followed through the same combination of sources, and not every analytical question requires continuous courtroom observation. As such, the Guide uses two monitoring modalities:

- **Full monitoring:** Combining hearing observation with review of legal acts and continuous case reporting; and
- **Tier 2 monitoring:** Relies primarily on authorised written sources and excludes systematic observation of hearings.

The matrix below helps you decide on a modality based on distinct analytical needs.

01 <i>Analytical need</i>	02 <i>Analytical need</i>	03 <i>Analytical need</i>
Courtroom management and conduct	Hearing productivity and stated reasons for adjournment	Oral participation and courtroom treatment
Full monitoring	Full monitoring	Full monitoring
Required	Preferred	Required
Tier 2 monitoring	Tier 2 monitoring	Tier 2 monitoring
Cannot establish from documents alone	Limited to sufficiently detailed official records	Cannot establish from documents alone



The modality may change as the case progresses. A Tier 2 case may later become suitable for full monitoring if its diagnostic importance increases and hearing access is available. A fully monitored case may later be followed primarily through judicial acts if the relevant question moves to appeal or another document-centred stage. Record the reason, effective date, approver and the effect on evidentiary continuity. Do not treat documents collected after a downgrade as equivalent to earlier direct observation.

Tier 2 broadens documentary coverage by increasing the volume of cases, thereby potentially addressing a defined evidence need, but it remains a controlled complement to direct trial monitoring. Controls should be established to ensure that Tier 2 document-only monitoring does not displace the methodology's observational core.

As Tier 2 is document-based, it consists primarily of indictments, procedural decisions, judgements, appeal decisions, official transcripts or hearing records, public registers, statistics, and other authorized written sources relevant to the defined inquiry.

6.2 Selecting cases for Tier 2

Use Tier 2 only to address a documented diagnostic need. The request should identify the unresolved question, specify the required documents and variables, state the expected analytical value, and identify the resource cost and the approving authority. It should also be made clear why the case was not selected for full monitoring.

Stakeholders may identify possible cases, but they must not control the sample. This requires a strict written protocol that includes every suggestion and the acceptance or rejection reason. There should also be a regular review as to whether stakeholder-suggested cases are having a disproportionate impact on the monitoring scope. To protect actual and perceived independence and objectivity, record stakeholder suggestions and the resulting selection decision internally. Protect the identity of the suggesting stakeholder or other sensitive details where confidentiality, security or access rules require it, but do not make the selection rationale methodologically invisible: the governing selection rule and any material departure from it should remain auditable and, where appropriate, reportable at an aggregate or methodological level. Tier 2 cases must remain within the approved field unless the Scope protocol is formally amended. Examples include:

01

Tier 2 category:

Thematically relevant cases

Diagnostic purpose

Examine an approved issue that is not sufficiently represented in the full-monitoring pool.

02

Tier 2 category:

Procedurally linked cases

Diagnostic purpose

Review related defendants, criminal networks, transactions or proceedings connected to a fully monitored case.

03

Tier 2 category:

Legal-classification pattern sample

Diagnostic purpose

Compare classification and charging patterns under a documented legal framework and comparable factual categories.

04

Tier 2 category:

Legal-act sample

Diagnostic purpose

Compare indictments, judgments, plea agreements, sentencing or confiscation reasoning across a wider documentary pool.

05

Tier 2 category:

Later-stage document follow-up

Diagnostic purpose

Follow appeal, retrial, enforcement or other developments where the defined question is visible in written acts.

06

Tier 2 category:

Recommendation-tracking sample⁴

Diagnostic purpose

When the engine re-enters Scope after Recommend or Track, test a defined implementation milestone across an approved documentary sample.

⁴ The policy engine is recursive rather than linear. Scope may be reopened whenever a later Step generates a new evidentiary need. Recommend may identify an implementation milestone that requires verification; Track may reveal that the existing evidence base cannot show whether practice has changed; Analyse or Explain may expose a gap that requires a narrower sample. In each case, the practitioner re-enters Scope for a defined purpose, applies the same disciplines of eligibility, selection, observation window and monitoring modality and depth, and constructs a proportionate new evidence stream.

Because Tier 2 monitoring bypasses direct courtroom observation, strict methodological safeguards govern the selection and analysis of these cases to protect the integrity of the findings. Such safeguards include labelling each Tier 2 observation by source type, source quality, and source independence. Exclude a case from a particular analysis if the necessary feature is absent or can be inferred only from an inadequate record. In addition, findings drawn from Tier 2 cases are strictly bounded by what can reasonably be inferred from the official written record. As such, drawing conclusions about courtroom dynamics, oral participation or observable treatment of parties from Tier 2 cases is prohibited where those matters are not reliably recorded. Equality-of-arms issues that are visible in written records, for example disclosure, access to the case file, time to respond or reasoned procedural rulings, may be assessed only to the extent those records support the claim. To prevent data conflation, these cases are clearly labelled in all datasets and reporting, ensuring that end-users can explicitly distinguish document-based observations from full hearing-based monitoring.

Finally, neither modality nor monitoring depth should become a hidden proxy for case importance. Full monitoring is not an award reserved for the most politically visible cases, and Tier 2 is not a residual basket for cases of little value. Each decision should reflect the evidence needed to answer the analytical question.

6.3 Confirm feasibility and access assumptions

Before finalising the observation design, determine whether the required evidence can be obtained within the mandate, available resources and project timeframe. This applies regardless of the selected monitoring modality. This is a critical interplay between Scope and Collect, as the latter must rely on the former to carry monitoring forward.

Prepare for instances when access or other relied-upon entry points or data become unavailable. For example, if published judgements are needed to identify confiscation decisions, create a plan for what happens when publication is incomplete or unavailable. Such changes should be controlled for and explained, rather than simply resulting in a reduction in rigour or approach.

7 Write down the evidence strategy

The evidence strategy is the bridge between Scope and Collect. It records what has been determined as the focus of the monitoring exercise and translates that decision into instructions for the collection process. As the methodological account of how the evidence base will be constructed, it connects the reform question to the case pool, the case pool to the observation window, and the observation window to the sources that must be secured. The matrix below provides an example of how the evidence strategy should be built and the use of each component.

01 Component:

Reform question

What the evidence strategy records:

The neutral question, applicable standard and the bounded claims the evidence base is intended to illuminate.

02 Component:

Baseline

What the evidence strategy records:

What existing information suggests, what remains uncertain, and which assumptions monitoring must test rather than inherit.

03 Component:

Subject-matter boundary

What the evidence strategy records:

The eligible offence types and the applicable inclusion, exclusion and classification rules.

04 Component:

Selection protocol

What the evidence strategy records:

The criteria, weights, thresholds and any justified exceptions used to rank eligible cases.

05 Component:

Case pool

What the evidence strategy records:

The initial list or category of cases selected for monitoring, together with the basis for inclusion.

06 Component:

Observation window

What the evidence strategy records:

The preferred entry point, intended end point, stages to be followed and known gaps in the procedural trajectory.

07 Component:

Monitoring modality and depth

What the evidence strategy records:

Whether each case receives full or Tier 2 monitoring, why that modality was selected, and what coverage and source intensity are required within it.

08 Component:

Source requirements

What the evidence strategy records:

The hearings, legal acts, registers, statistics, contextual material and other sources that Collect should seek.

09 Component:

Access plan

What the evidence strategy records:

Where the relevant material is expected to be obtained, which access channels will be used and what restrictions are anticipated.

Note: Access, collection variables and evidence gaps are first identified in the context-baseline task and are narrowed here to the selected case pool and evidence strategy.

10 Component:

Collection variables

What the evidence strategy records:

The principal facts, events, decisions and attributes that must be recorded consistently across the pool. This is already identified in the baseline task but now narrowed to the case pool.

11 Component:

Evidence gaps

What the evidence strategy records:

Material that is unavailable, inaccessible or not directly observable, and the effect this may have on later findings. This is already identified in the baseline task but now narrowed to the case pool.

12 Component:

Review and change points

What the evidence strategy records:

When the scope, case list, observation window or monitoring modality should be reconsidered.

7.1 Dealing with assumptions

Every evidence strategy contains assumptions. The practitioner may assume that sufficient cases will become available, that hearings will remain public, that indictments can be secured, that offence classifications are comparable, that a particular procedural stage will reveal the relevant practice, or that Tier 2 records will contain sufficient reasoning to support cross-case analysis. Assumptions should be recorded and continually assessed, both for their impact on the ongoing monitoring exercise and for how they might affect later analysis. Ensuring the visibility of assumptions allows for revisions to the approach without assuming infallibility.

7.2 Hand the Scope, with the evidence strategy, to Collect

Once the full evidence strategy is determined, the process can move to Collect. Before doing so, the following must be clear:

- *Can the reform question be stated in a single sentence? Are the boundaries, units and intended claims clearly defined? Can another practitioner apply the eligibility and priority rules and reconstruct every inclusion, exclusion and monitoring-modality decision?*
- *Are assumptions, access constraints, ethical and safety risks, evidence gaps, and review triggers recorded? Does the handover tell Collect what to obtain and tell Analyse what the evidence cannot support?*

Based on this, Collect receives the approved selection protocol, details of decisions, monitoring-modality assignments, observation windows, access and feasibility assessments, and a review schedule. Collect will necessarily need to test some of the Scope assumptions, noting that a return to Scope may be necessary if case flow, access, legal classification, safety, source quality, or new evidence materially changes. That said, operational inconvenience alone does not justify undocumented re-scoping.

► Before you move on



ETHICAL FRAME NOTE

Scope fixes the ethical basis for everything that follows, because every later Step inherits the field, the pool and the modality decided here. Two principles are particularly relevant at this Step:

Agreement. Systematic trial monitoring depends on programme-level agreement and institutional cooperation for reliable schedules, records and access to counterparts. This is distinct from the public's right to attend open proceedings where domestic law provides it, and an access arrangement must never create an institutional veto over case selection, findings or lawful observation. Cooperation may inform the pool; it must not control it, whatever the modality. Selection follows the stated criteria and records every exclusion.

Do no harm. Selection puts people under observation who did not choose it. Assess whether monitoring, recording or storing information could expose victims, witnesses, journalists, justice actors or practitioners, and apply the applicable data-protection and anonymisation rules. Safety or rights risks may justify exclusion, delayed inclusion or a less intrusive modality. Record the reason, not only the decision.



COMMUNICATION CARD

At this Step explain why the inquiry exists and what institutional value it is intended to serve, rather than introducing it through the deficiencies it might uncover. What you can say starts with the mandate, the reform question, the institutional and temporal scope, the types of practice and evidence being examined, and the principal safeguards. Be equally clear about what the inquiry is not designed to do: for instance, it examines professional and institutional practice, it does not review the merits of a decision or intervene in a pending case. What you get back from the communication is information about legal changes, institutional reorganizations, parallel reforms, access constraints, and distinctions between institutions that would change what must be examined. In that sense, a counterpart may show that the period, actor or source you intended to examine is methodologically inappropriate — and that is a reason to adjust. Discomfort with the subject is not. If the information reveals a genuine reason to change, take it. back to Scope and document it there.



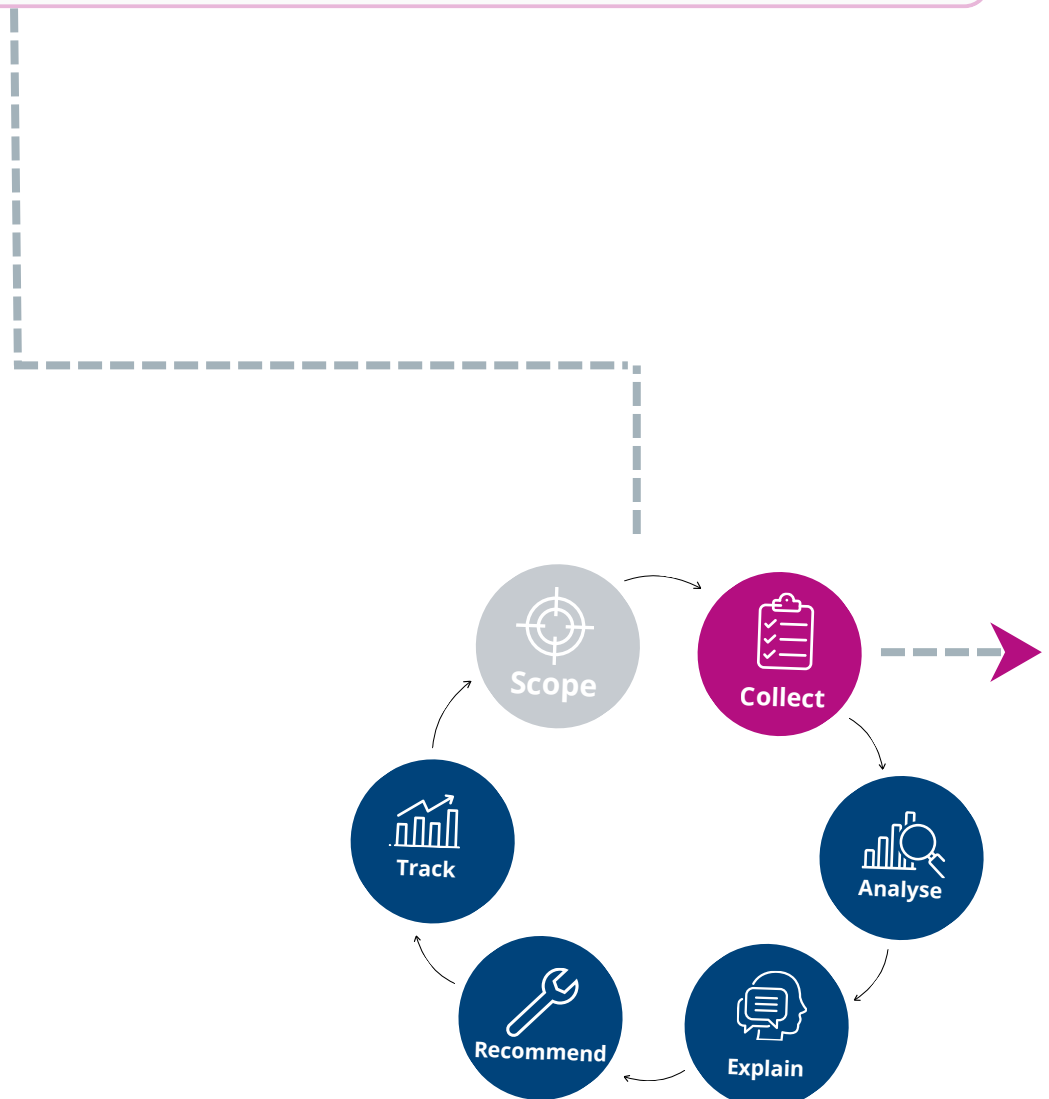
ADAPTATION NOTE

Within adjudicative monitoring, the practitioner may change the case typology while retaining the same scoping sequence. The selection criteria must be re-derived from the new reform question; OCC-specific status, gravity and media-attention weights do not transfer automatically. Although the scoping sequence still applies, eligibility, priority and observation determinations should be derived from the reform question itself. For example, a project monitoring gender-based violence cases might prioritize victim vulnerability, use of protective measures, repeat offending, institutional co-ordination, evidentiary sensitivity and the risk of secondary victimization. Adapting to a different justice -system function or institutional site is a significant methodological change. Investigation, legal aid, enforcement, policing, regulation or legislative processes require equivalent observation points, records, access rules and safeguards. This Guide is primarily intended for adjudicative monitoring.



STEP-CHECKLIST

- ☐ Can the **reform question** be stated in one sentence, and does it name a problem rather than a predetermined answer?
- ☐ Is the subject-matter **boundary** stated precisely enough that a stranger could apply it to a new case without asking you?
- ☐ Does **every case in the pool** carry a stated reason – priority score or documented diagnostic basis – for its inclusion, together with its monitoring modality?
- ☐ Have **assumptions** been written down as assumptions, not folded silently into the criteria?
- ☐ Have anticipated **evidence gaps** been logged, so Collect encounters them as known limits rather than surprises?



Step 2



Collect

► Where you are in the cycle

Scope determines which practice area will be examined, which cases or proceedings will enter the evidence base, and how deeply each will be examined. Collect turns those decisions into a structured, source-tagged body of material. It does not yet produce validated findings: Analyse determines what the collected material can support. Because the same evidence base then supports explanation, recommendation development and later tracking, it must be collected rigorously.



► Why this Step is important

Sound analysis and reform design require source material that is relevant, traceable, comparable and collected under clear limits. Collection is therefore the first point at which the evidence base is built.

► What happens if we are not intentional about it

Without intentional collection, the evidence base becomes uneven. Some cases are richly documented, while others are sparsely documented. Observations from hearings are not comparable. Prosecutorial and judicial acts are reviewed inconsistently. Contextual material is mixed with directly observed or documentary evidence. Statistics are treated as self-explanatory, and stakeholder views are mistaken for procedural facts. If source, date, collector, collection modality, access status, version, language or translation status, and restrictions are not recorded, the evidence cannot be traced or independently reviewed.

► What you produce

By the end of this Step, the practitioner should have:

a confirmed, continuously updated monitored-case list; an access map; hearing records for every monitored session; a legal-act collection and coding log; official statistics and administrative data, where available; consultation and media logs; an evidence-gap register; a data-management and security plan; a coding and quality-assurance record; a consolidated, source-tagged evidence base; and a short handover memorandum to Analyse, stating coverage, limitations and unresolved gaps.

THE STEP AT A GLANCE

1

Establish the list of cases to be monitored

Which cases from the eligible field will be monitored?

Turn the evidence strategy made in Scope into a concrete and continuously updated case list.

2

Secure continuing access

How will data become available as each case develops?

Establish reliable access to hearing schedules, public hearings, legal acts, official records and institutional information.

3

Monitor hearings and maintain the case record

Monitor hearings and maintain the case record?

Record each procedural event and emerging issues across the life of the case.

4

Collect and review key legal acts

How were charges, facts, evidence, law and decisions formally framed and reasoned?

Collect the key judicial and prosecutorial acts and review them through consistent qualitative criteria.

5

Complete and bound the evidence base

What other sources are needed?

Add the complementary source streams required to situate, test and qualify the core trial-monitoring evidence: official statistics, consultations, media monitoring, and evidence gaps.

6

Assemble an evidence base

What can be analysed, and on what evidentiary terms?

Consolidate the material into a structured, source-tagged evidence base that shows what was directly observed, what comes from legal acts, what is statistical or contextual, what derives from professional perspectives, and what remains unavailable or unverified.

► What Collect does: from evidence strategy to evidence base

After Scope defines what the monitoring exercise needs to observe, Collect creates the list of cases to monitor, secures the required sources, creates indexed, source-tagged records, and documents uneven coverage and unresolved gaps. Analyse should receive only material that has passed the collection quality-assurance threshold, together with a coverage statement highlighting missing or uneven evidence.

The policy engine's evidence base is more than a repository of documents and hearing notes. It is the structured, source-tagged record of practice that enables Analyse to distinguish a procedural event from a pattern, Explain to test possible drivers, Recommend to justify a reform recommendation, and Track to verify whether practice has changed.

But between an evidence strategy and an evidence base lies a sequence of practical decisions: identifying real cases that meet the criteria, securing access to hearings and legal acts, recording proceedings consistently, collecting complementary sources, and making visible what could not be obtained or verified.

For that reason, collection must produce evidence that is:

- relevant to the reform question;
- traceable to its source;
- recorded consistently across cases;
- comparable where comparison is methodologically sound;
- clearly distinguished by monitoring modality;
- honest about its gaps and limits.

The Step begins with two related yet distinct tasks, both of which require access to case information and to judicial institutions:

- Establish the list of cases to be monitored: Determine which proceedings meet the approved Scope criteria; identify the actual proceedings; verify their offence classification, status and procedural stage; and decide whether they satisfy the Scope criteria. This task requires access to case information.
- Monitor the confirmed cases: To undertake monitoring, there must be access to hearings as they occur, to legal acts as they are issued, and to the official and contextual material required to track the case's trajectory.

Keeping these tasks separate prevents access arrangements from determining the monitored pool, which would dilute the established criteria.

1 Establish the list of cases to be monitored

The first collection task is to turn the selection logic established in Scope into a concrete and defensible list of monitored cases.

Scope defines eligibility, priority, timing, monitoring modality and monitoring depth. It does not, by itself, produce the case list. Between a set of criteria and a monitored pool lies a practical task: establishing which relevant proceedings exist, where they are handled, how they are legally classified, what procedural stage they have reached, and whether sufficient information is available to apply the agreed criteria.

The practitioner should therefore maintain two linked records:

A potential-case register, containing proceedings that may fall within the scope; and

A confirmed monitored-case list, containing cases that have been checked against the evidence strategy and assigned a monitoring modality.

The distinction matters. A case may appear relevant in media reporting or stakeholder information yet not relate to the selected offence. It may meet the subject-matter criteria but be at a procedural stage that offers limited observational value. It may rank highly yet remain inaccessible. Alternatively, it may initially appear low priority but later acquire diagnostic relevance through a linked proceeding, a new indictment, or an emerging need for evidence elsewhere in the policy engine.

1.1 Identifying potential cases

Information about potentially relevant proceedings may come from any reliable, context-appropriate source. Common channels include judicial and prosecutorial institutions, the media, and civil society and other informed actors. Although these sources are detailed further below, this should be treated as an open-source set, not an exhaustive hierarchy. Information from any channel must be verified against authoritative legal records before a case is confirmed. This may be:

The institutions themselves are the primary source. Judicial and prosecutorial councils, court presidents and heads of prosecution offices can supply lists of ongoing cases with the precise legal qualification of the offence and the status of the accused.



Media – especially outlets specialising in investigative reporting.

Civil society organizations, particularly those supporting whistleblowers, know of proceedings before they are listed anywhere.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM project, summarized.

These channels should be used complementarily, not interchangeably. Media and civil society can broaden visibility by identifying proceedings that may not yet appear on institutional lists, while institutional sources provide the primary point of verification for case identity, legal qualification, procedural status, and other formal case information. When underlying judicial or prosecutorial acts are required, obtain them through official channels to ensure source provenance is clear and the evidentiary record remains authoritative. At the same time, the practitioner may use alternative channels to identify possible cases and contextual information. Using and interacting with each of these sources requires a specific approach (for more details on the approach we recommend, consult the ODIHR TM Manual (2012)).

1.2 Apply the Scope criteria in sequence

Each potential case should be checked against the factors already recorded in the evidence strategy.

The practitioner asks the following questions in order:

Eligibility: Does the charged offence fall within the subject-matter boundary?

Priority: How does the case score against the agreed substantive criteria?

Timing: Is the case within the observation window, or can its earlier trajectory be reconstructed sufficiently?

Modality: Does the case require full monitoring or Tier 2 document-based review?

Depth: Within that modality, what coverage and source intensity are required to support the intended claim?

Feasibility: Can the necessary sources realistically be accessed?

Contribution to the pool: Does the case add thematic, geographic, procedural or comparative value to the monitored sample?

This sequence prevents a visible or politically prominent case from circumventing the methodology. It also prevents feasibility from becoming the only selection rule.



Practitioner Note

A case enters the monitored pool because it meets the approved criteria, regardless of whether the institution processing it flags it. Institutional co-operation is essential for access, but it must not create a veto over the initial pool, Tier 2 additions, stakeholder referrals, or any later targeted sample. Document and authorize or review every inclusion, exclusion, and monitoring-modality change in accordance with the approved selection protocol and the organization's governance arrangements, so independence and objectivity can be demonstrated

a) Geographical spread

Where relevant to the reform question, the practitioner should consider the territorial distribution of the monitored pool. A pool drawn entirely from the capital city supports claims about the institutions and proceedings represented there, not necessarily about all judicial bodies. Where certain cases cluster in specialized or central courts because of statutory jurisdiction, prosecutorial centralization, or institutional design, the practitioner should record that concentration rather than artificially correct for it. Generally, the practitioner should explain the geographic spread the pool can and cannot support.

b) Sample size and the case pipeline

The inflow of eligible cases is not controlled by monitoring. In some jurisdictions, only a small number of new high-priority proceedings may arise each year. The evidence strategy may set a target range and allow medium-priority cases to enter when too few high-priority cases are available. When the pipeline remains thin, Tier 2 review may broaden documentary coverage, but it does not replace the observational depth of hearing monitoring.

Small case numbers constrain claims about prevalence, frequency and jurisdiction-wide trends, but they do not automatically weaken every finding.

Serious and complex OCC cases are selected in part because they provide a diagnostically valuable stress test of institutional performance: they place exceptional demands on legal reasoning, evidentiary handling, co-ordination, resources and procedural management, often under heightened public scrutiny. A small number of deeply observed cases may therefore support strong case-level or cross-case analytical findings where the evidence is sufficiently rich and triangulated, and the claim is appropriately bounded. The strength of a finding depends on the fit between the evidence and the level of the claim, not on case numbers alone.

c) Forward projections

Practitioners should gather information on expected high- and medium-priority proceedings to plan resources, anticipate hearing workloads and sequence case intake. These projections do not equate to confirmed cases, but they enable practitioners to plan resources and contingencies. Projections remain programme-planning information and do not enter the monitored evidence base until the case has been verified and confirmed.

1.3 From potential case to confirmed monitored case

A case moves from the potential-case register to the confirmed list once the practitioner has enough verified information to determine that:

- it falls within the subject-matter field;
- its priority category has been assessed;
- its procedural stage is suitable;
- its monitoring modality has been assigned;
- the minimum necessary access appears feasible.

Confirming the list indicates which proceedings will be monitored. The next Step sets out how the practitioner will access the evidence those proceedings generate as they unfold.

1.4 Continue to build and adjust the monitored pool

The case list should not be treated as settled at the outset, or even the close, of Collect. New indictments are filed; proceedings move to trial, appeal or retrial; cases are severed or joined; access conditions change; and later analysis or tracking during Analyse or Track may identify an evidence gap that the current pool cannot address. Case identification and confirmation should therefore operate on a continuous, rolling basis, in accordance with the requisite processes.

2 Secure continuing access

With the list confirmed, the process and the need for access shift to the evidence itself, and its object accordingly changes. What is now required is, first, the schedule of hearings in the selected cases, so that relevant sessions are neither missed nor learned of too late to attend, and second, the legal acts as they are issued.

Access is not a one-off permission. It is an operating arrangement that must function throughout the monitoring period. The practitioner may be aware of the proceedings but lack information about the hearing schedule or may be able to attend hearings but not receive legal acts. Continuous access therefore requires more than an initial institutional agreement. It requires a practical information system.

Continuing access depends on practical, complementary, and ongoing arrangements rather than a one-off permission. These may include designated focal points, a regular docket-review schedule, and multiple routes to the same information. A memorandum of understanding may be useful where an institution prefers a written basis, but it is not a precondition if monitoring already rests on established co-operation. Any formal or informal arrangement should preserve independent case selection and findings, ensure equal and non-discriminatory access, clarify the distinction between public and non-public material, ensure continuity despite staff turnover, and include a process for refusal, dispute, or termination. It should not require any quid pro quo or compensation, including sharing any findings. For further guidance on trial-monitoring access arrangements, consult the ODIHR TM Manual (2012).



Practitioner Note

Securing an access channel does not mean that the evidence base is complete. An institution may formally agree to co-operate but provide schedules inconsistently. A public database may exist but omit key acts. A court may permit hearing attendance but not provide written decisions. A focal point may respond regularly for some cases and not others.

The practitioner should therefore distinguish:

- Formal access – Permission or institutional agreement exists;
- Operational access – Information and material are received in time to be useful;
- Monitoring-sufficient access – The source set required for the defined monitoring purpose is available across the relevant case trajectory.

This distinction prevents an access arrangement from being reported as successful merely because it exists on paper. Access challenges should also be assessed as part of the exercise because they may provide relevant information about institutional practice; they should not be treated only as an operational inconvenience.

Access is a prerequisite for collection, not evidence of completeness. Observing hearings and reviewing key prosecutorial and judicial acts form the core trial-monitoring streams. Official statistics and administrative data provide distributional context, while consultations and media materials provide professional, reported, or contextual information. Evidence-gap records indicate what remains unavailable or unverified. These streams provide different forms of information and work together to form the evidence base.

3 Monitor hearings and maintain the case record

3.1 Which hearings

Hearing observation constitutes the core method of data collection in trial monitoring and is the activity that most clearly distinguishes it as a specific type of rule of law intervention. It involves the systematic observation of court proceedings in the cases selected for monitoring, with a view to capturing how the justice process operates in practice. It is also the source that, in principle, should be accessible in any trial monitoring project given the public nature of criminal trials. It should be noted, however, that national criminal procedures and international standards foresee the exclusion of the public from hearings in specified circumstances. While attendance at closed hearings can be important to monitoring programmes and can be negotiated with authorities, this Guide's approach to data collection limits monitoring to public hearings because of the matter's sensitivity.

This also entails that cases are monitored starting from the confirmation (or equivalent) of the indictment until their completion; accordingly, the pre-trial and investigation stage, insofar as they are not open to the public, are not monitored.

SOURCE



This said, the following types of hearings should be monitored as a minimum (to the extent envisaged under the domestic codes of criminal procedure):

- Hearing on confirmation of the indictment
- Plea hearing
- Hearing on confirmation of plea agreement
- Status conference hearings
- Main trial hearings
- Sentencing hearings
- Appeal hearings
- Hearings on imposition or review

As a rule, hearings are monitored in person. Exceptionally, where attendance is not possible, one may seek access to alternative sources, such as audio recordings, transcripts, or official minutes, in order to ensure continuity and completeness of the record. In such instances, the use of these sources is explicitly indicated in internal documentation.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM project, summarized.

3.2 Modalities of recording data

Monitoring practice generally requires extensive note-taking during hearings, often in a form that closely mirrors the exchanges in court. At the same time, given the breadth of data arising during most hearings, not everything can or should be recorded, and taking verbatim notes may be impossible. Courts may also prohibit recording devices. Clear methodological guidance is essential to ensure the consistency and relevance of the data collected, including a protocol for correcting notes after the proceedings.

Two distinct tools used at this stage enable the collection and later analysis of hearing data.

a) Daily Hearing tool

The first is the Daily Hearing tool. It records the hearing as an event but is neither a transcript nor an analysis. Its purpose is to preserve the procedural event in sufficient detail by selecting the relevant information needed to support later analysis of how the proceeding was conducted.

This tool records the hearing as an event. It should capture, at minimum:

01 <i>Recording area</i> Identification <i>What should be preserved</i> Case identifier, date, court, panel, procedural stage and hearing type.	02 <i>Recording area</i> Access and publicity <i>What should be preserved</i> Whether the hearing was public, whether the practitioner attended in person and whether any restrictions applied.	03 <i>Recording area</i> Participation <i>What should be preserved</i> Which parties, counsel, witnesses, experts and other relevant participants were present or absent.
---	---	---

04 <i>Recording area</i> Procedural activity <i>What should be preserved</i> Short summary of type and content of presented evidence (witnesses, documents, etc.), motions raised, objections made, procedural questions addressed and decisions issued.	05 <i>Recording area</i> Hearing productivity <i>What should be preserved</i> Whether and how the hearing advanced the proceeding.	06 <i>Recording area</i> Delay or disruption <i>What should be preserved</i> Adjournments, interruptions, missing participants, incomplete expert reports, technical problems or other procedural events affecting progress.
07 <i>Recording area</i> Procedural safeguards <i>What should be preserved</i> Observable issues concerning fair trial and human rights safeguards, such as participation, equality of arms, witness protection, impartiality or other relevant guarantees.	08 <i>Recording area</i> Source limits <i>What should be preserved</i> What could not be heard, observed, obtained or verified.	09 <i>Recording area</i> Next Step <i>What should be preserved</i> The next scheduled hearing or procedural development, where known.

b) Case Report tool

The second is the Case Report tool, which builds the case map and serves as the case's first-order analytical record. It records the basic case information, procedural chronology, defendant-level data, media context, hearing summaries, and emerging issues or analytical leads.

It is the first stage in which monitoring observations are organized into structured, provisional analytical leads through the lenses of Capacity, Fairness, and Efficiency. It moves beyond a purely chronological record by identifying issues or possible concerns that later analysis may need to test, while preserving the underlying observations and sources. It does not convert source material into validated findings: Analyse assesses whether those leads are supported as evidence for a claim against the applicable standard and at the appropriate level. Most importantly, it makes clear that findings are not to be analysed in isolation at the level of a single hearing. They are case-related and must be updated throughout the proceeding, with references to the specific hearing or source where the issue arose. As such, the Case Report remains case-specific and does not claim broader institutional causes.



Practitioner Note

The subsections that follow, within this Step as well as section 3.2 in Step 3, lay out the core analytical content and functions that a Case Report should capture. Unlike the Daily Hearing tool, the Case Report is not prescribed here as a single standard form. Its structure should remain adaptable to the legal context, case type, monitoring scope and analytical needs of the exercise. What should remain constant is its function: to connect evidence across the life of the case, preserve source references and organize the material so that case-level and later cross-case analysis can be conducted systematically.

A Case Report should include guiding questions organized under the headings of capacity, fairness and efficiency. This does not mean that the practitioner completes the analytical work for Step 3 while collecting evidence.

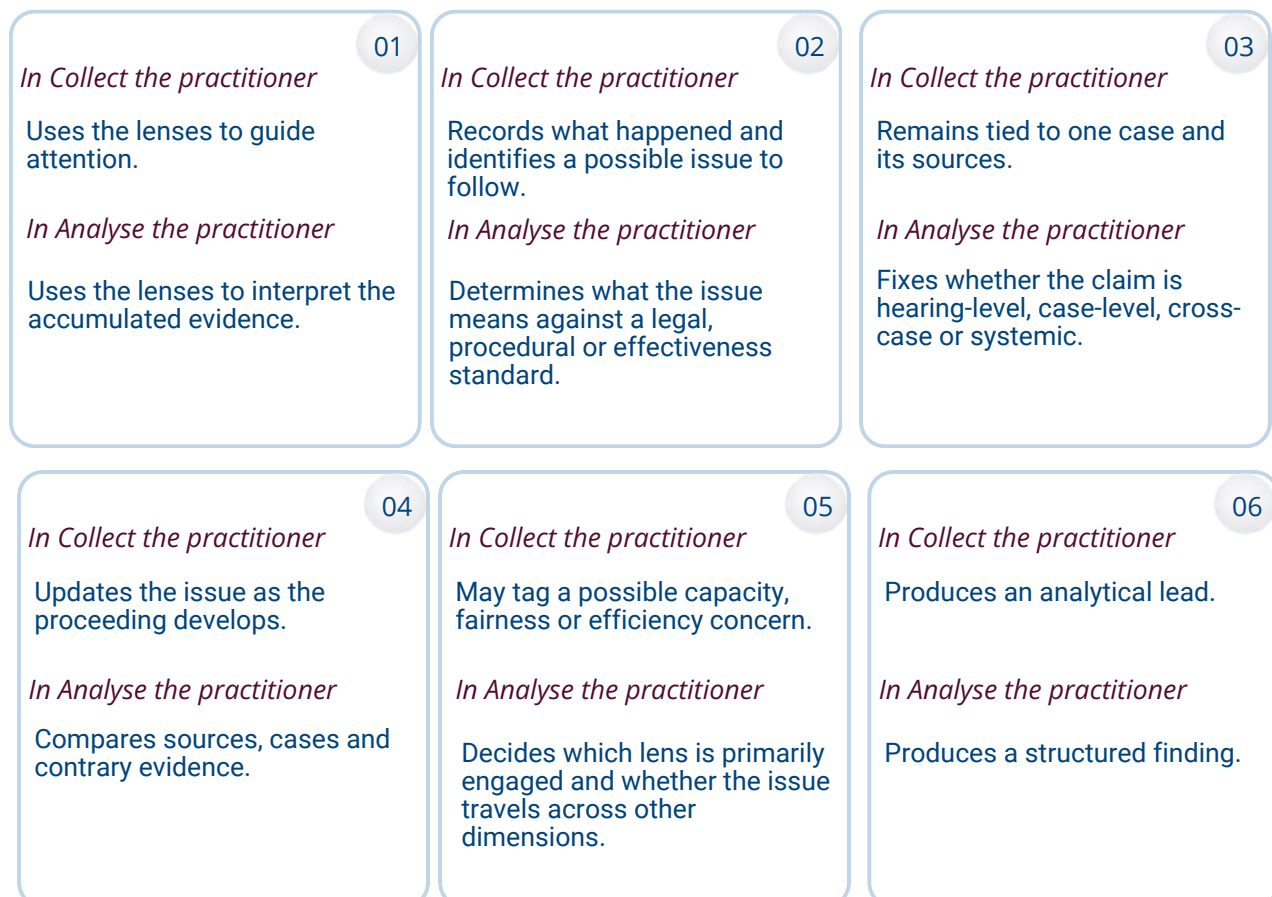
The questions serve a more immediate purpose. They sensitize the practitioner to practices that may later prove analytically important and ensure that relevant evidence is consistently recorded during the case. Without such prompts, practitioners may record the formal sequence of a hearing but overlook the conduct, decisions or recurring procedural conditions that later matter for assessing effectiveness.

For example, the Case Report may prompt the practitioner to ask:

1. Was the prosecution's case theory presented coherently?
2. Did the court manage interruptions or witness intimidation?
3. Were hearings repeatedly adjourned for the same reason?
4. Were procedural decisions visibly reasoned?
5. Were available protective or financial-investigation measures used?
6. Was the hearing productive, and what does the record show about party preparedness?

At Collect, these are evidence-seeking questions. They tell the practitioner what facts, conduct, statements and written sources should be preserved. They do not yet authorize or prompt analytical conclusions or hypotheses.

The distinction in the use of the Case Report is one of function and evidentiary burden. The same guiding question may therefore appear in both Steps, but its role changes.



The Case Report is therefore more than a data-entry tool, but less than the final analysis. It is a first-order analytical organizing tool. As such, it performs four functions:

It directs attention: The guiding questions help practitioners recognize potentially relevant conduct, decisions and good practices while they are still observable.

It preserves the source: Every concern or good practice is tied to the hearing, legal act, statement or other source from which it arose.

It creates an issue line: A possible concern is followed across the case rather than judged from one hearing in isolation.

It prepares the hand-off to Analyse: By the end of the proceeding, the Case Report shows what recurred, what was resolved, what was contradicted, and which questions remain open.

This is why the Case Report questions are organized through the same three lenses later unpacked in Analyse. The collection instrument and the analytical framework must speak the same language. Otherwise, Analyse may ask questions that Collect was never designed to answer.

4 Collect and review key legal acts

4.1 Which legal acts

The types of acts to be collected and reviewed are identified on the basis of their significance as key milestones in criminal proceedings. While the precise range may vary depending on the procedural framework of each jurisdiction, it generally includes, at a minimum: indictments; decisions on the confirmation or rejection of indictments; first-instance verdicts; appellate decisions (on both verdict and sentence); retrial judgements; final decisions on a case; and decisions related to plea agreements.

In addition, the Project may seek access to other acts where necessary to examine issues of particular relevance to OCC cases, provided that such documents are not subject to confidentiality restrictions. These may include, for example, decisions on special investigative measures, asset seizure, witness protection, the admissibility of evidence, and mutual legal assistance. Best practices for accessing judicial acts are set out in detail in the ODIHR Trial Monitoring Manual (2012).

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

SOURCE



4.2 Modalities of recording data

Legal acts are reviewed and coded using qualitative criteria and, where useful, structured variables.

Two distinct tools support legal-act evidence collection and prepare the material for later analysis; the legal acts review tool and content analysis register.

The Legal Acts Review tool applies a consistent structure to the indictment, first-instance judgement and appeal decision, with adaptations for other acts. Each review opens with the act's identifying details and case categorization, sets out a plain summary of what the act charged, found or decided, and then subjects it to a fixed grid of reasoning criteria. Each category is reviewed against criteria focused on form, clarity, reasoning, and the link between facts, evidence, and law. The purpose is not to second-guess the outcome of a case, but to assess whether the written act allows the reader to understand how the charge was framed, how facts were established, how legal issues were resolved, and how the conclusion was justified. The reviewer may record omissions, internal contradictions, or a failure to connect facts, evidence and law where the applicable legal standard permits such assessment.

The indictment is reviewed against the applicable domestic procedural requirements and the elements of the charged offences. The review asks whether the indictment is clear and formally complete; whether the alleged facts align with the material and mental elements of the offence; whether relevant modes of liability, and where applicable, corporate liability, are stated; where domestic law requires an evidentiary basis or linkage to be set out, whether that material is organized in relation to the alleged facts and charges; and whether any confiscation proposal is grounded in the applicable legal basis and the financial or asset evidence identified in the act. The review assesses the construction and legal sufficiency of the charge under domestic law; it does not determine whether guilt will ultimately be proved.

The judgement is assessed first against binding domestic law and applicable human-rights standards. Where applicable, for example, Consultative Council of European Judges (CCJE) Opinion No. 11 may then serve as a non-binding European benchmark for clarity and reasoning quality; it does not create separate legal obligations. The review asks, as applicable, whether the act complies with formal requirements; identifies the material factual and legal issues; addresses material submissions to the extent required by the applicable standard; links factual conclusions to the cited evidence where the applicable standard requires such reasoning; explains the applicable law; remains internally coherent and free of material contradiction; and provides reasons sufficient under the applicable standard. Where domestic law requires reasons on mitigating or aggravating factors, sentencing analysis should examine how the decision addresses their relevance to the outcome rather than merely noting their existence.

For appeals, the tool applies comparable criteria after first identifying the appellate court's powers and duties under domestic law. Where the first-instance decision is quashed or remitted, the review should assess whether the appellate decision identifies the relevant error and provides the guidance required by that legal framework.

This tool therefore serves two purposes. First, it captures the written act as a source of evidence. Second, it records selected features of that act against predefined categories to support subsequent analysis of capacity, fairness and efficiency.⁵ For example, where the applicable domestic framework requires or permits assessment of these features, an indictment may leave unclear the connection between the evidentiary basis, the factual allegations and the legal elements of the charges, or the alleged role of individual defendants. A judgement may not explain how conflicting evidence was assessed, may leave material arguments raised by the parties unanswered, or may provide insufficient reasoning for a particular legal conclusion or sanction. An appellate decision may identify specific errors in the first-instance proceedings and record guidance relevant to the handling of the case on retrial.

For larger document samples, the practitioner may also use a separate Content-Analysis register.⁶ Unlike the judicial-act review tool, which assesses each act in depth, the register records selected features across many documents using the same predefined categories. This makes it possible to compare the documents systematically, identify recurring patterns and examine how particular forms of reasoning, language or legal practice are distributed across the sample.

⁵The predefined features and categories should reflect the purpose of the monitoring exercise. Its categories should be clearly defined, applied consistently and linked back to the underlying source. Where several practitioners use the system, they should align their approach and record any decision rules.

⁶Content analysis is the systematic examination of a defined set of documents through consistently applied categories. It allows selected qualitative features of the text to be converted into structured, countable data, while preserving the underlying passages needed for interpretation. This makes it possible to identify and compare recurring patterns, differences and distributions across the sample.



Method Note

Where structured coding is used, define each variable and its permitted values in a short codebook or coding protocol before aggregation. State the unit being coded (for example, hearing, case, defendant, charge, legal act or decision); distinguish missing or unknown information from “not applicable”; and link each coded value back to its source. Test ambiguous categories on a small set of records and record the decision rules. Where more than one practitioner codes material, calibrate the coding on common records and document material disagreements and their resolution. Version and date any change to a coding rule and assess whether earlier records need to be re-coded. The purpose is not statistical sophistication but reproducibility: another practitioner should be able to understand what a category means and how a value was assigned.

5 Complete and bound the evidence base (hearings and legal acts)

The core evidence is sufficiently complete when the defined source materials have been obtained, or when every missing item has been recorded and its effect on the inquiry assessed. Completeness is claim-specific: the material may be sufficient to assess the written reasoning in a judgement while remaining insufficient to assess courtroom participation, pre-trial decision-making, or the reason for an adjournment.

Before closing the core collection for a case, the practitioner should confirm that procedural details are current and complete; that the relevant records and acts have been collected; and that any missing material has been addressed. Collection may not result in an exhaustive record, but rather in evidence sufficient to assess the defined question.

In completing the evidence, the division of labour between the tools should remain clear:

The Daily Hearing tool records the event;

The Judicial-acts Review tool examines the formal legal product; and

The Case Report tool connects several elements across the life of the case.

For example, a hearing record may show that the defence challenged the link between the financial evidence and a particular charge. The judicial-act review may later show whether the judgement addressed that argument and linked the evidence to the relevant legal elements. Analyse then determines what that combined record can reasonably support.

This is why the Case Report serves as the bridge between hearing observation and judicial-act review. It ensures that the two core evidence streams are collected separately yet understood together. In this way, it enables the practitioner to determine whether an issue first observed in court was subsequently addressed in writing, whether the written reasoning aligns with the procedural pathway, and whether the same concern or good practice reappears at later stages.

5.1 Add the complementary source streams

Hearing observations and legal-act reviews constitute the core evidence of practice. They do not, on their own, provide the full context needed to understand the monitored case pool.

The practitioner must also understand how the cases fit within the system's broader caseload, how judicial and other actors experience the conditions surrounding the proceedings, how cases are portrayed in the public information environment, and where the evidence base remains incomplete.

The practitioner should then add the complementary source streams required to contextualize, test and qualify the core trial-monitoring evidence, including:

- consulting official statistics to understand volume, distribution and time;
- consulting judicial and other relevant actors to capture professional perspectives, practical constraints and possible explanations;
- monitoring media to identify cases and document the public and narrative environment;

These sources do not serve the same evidentiary function. Statistics can describe distribution within their defined population. Consultations can evidence reported perspectives and help illuminate the broader reform context. Media can document published narratives, public statements and contextual signals, including possible pressure. The objective is not to collect every available source, but to ensure the evidence base is sufficiently complete to assess the reform question and sufficiently bounded to prevent overclaiming.

a) Official statistics

Official statistics give the practitioner a quantifiable understanding of the operational environment.

Official statistics help understand whether monitored cases are exceptional, representative of a broader pattern, or part of a distribution the system itself can or cannot see.

SOURCE



ODIHR's methodology treats official statistics as an important source because the project monitors only the most serious OCC cases in each jurisdiction. It identifies basic statistics that should be available, including the number of indictments, first-instance verdicts, appeal decisions, final decisions, conviction/acquittal rates and average lengths of trial, appeal and overall proceedings. It also states that the absence or unreliability of such statistics should be noted as a structural shortcoming.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

Depending on the reform question and the available institutional data, relevant statistics may include:

- 1) numbers of indictments;
- 2) first-instance judgements;
- 3) appeal decisions;
- 4) final and binding decisions;
- 5) convictions, acquittals and discontinued proceedings;
- 6) plea agreements;

- 7) retrials;
- 8) sanctions imposed;
- 9) confiscation orders and amounts;
- 10) average or median duration of first-instance, appeal and overall proceedings;
- 11) numbers and reasons for adjournments;
- 12) caseloads by court, prosecution office, judge or prosecutor, where appropriate and lawful;
- 13) case distributions by offence type, procedural stage, geography or defendant category.

Statistics, in themselves, are not self-explanatory. They show volume, distribution and time, but not the quality of charges and evidence, the proportionality of sanctions, the fairness of proceedings, the reasons for delay, or whether reforms have changed practice. They are therefore collected as evidence within these limits and interpreted alongside observations, acts and the procedural history. Because the monitored pool is purposively selected, official statistics should be used to locate monitored patterns within the wider caseload or to test whether similar patterns are visible there, not to describe the monitored pool itself as statistically representative.

Before a statistic enters the evidence base, the practitioner should establish:

- the institution that produced it;
- the unit of analysis;
- the period covered;
- the population included and excluded;
- the offence classifications used;
- whether defendants, charges, cases or decisions are being counted;
- whether first instance and final outcomes are distinguished;
- whether the data are complete;
- whether the definition or collection method changed over time.

A conviction rate, for example, may refer to convicted defendants, charges resulting in conviction or cases with at least one conviction; all the above may refer to finalized cases exclusively or to all cases where at least a verdict was issued. These measures are not interchangeable.

The source record should therefore preserve both the figure and its definition. A statistic without its denominator, unit and coverage rule is not yet usable evidence.



Method Note

Where basic statistics on serious cases are unavailable, unreliable or not disaggregated in a way that allows meaningful analysis, that is not only a project inconvenience. It may indicate a system's limited ability to know, manage and explain its own performance. The absence of data should be recorded as an evidence gap and, where the underlying duty, institutional process and source fit support it, as a potential broader institutional issue for Analyse.

b) Consult with judicial and other relevant actors

Consultations with judicial actors and other professionals involved in the administration of justice (such as judges, prosecutors, defence counsel, and court staff) provide a valuable complementary data source for a trial monitoring project. As noted in the ODIHR Trial Monitoring: A Reference Manual for Practitioners (2012), engaging with actors directly or indirectly involved in the justice system can yield significant additional insight, both by supplementing and by testing issues identified through hearing observation and legal-act review. Such interactions may reveal aspects of the system's functioning that are not readily observable through observation alone.

They may help the practitioner understand:

- how formal procedures operate in practice;
- what resource or organizational constraints affect case handling;
- how responsibilities are divided between institutions;
- where legal rules are unclear or difficult to apply;
- why particular tools are used or avoided;
- how access, scheduling, expertise or co-ordination problems arise;
- whether an observed practice is unusual or widely recognized;
- which possible explanations require further evidence.

Consultations should be conducted under clear attribution and confidentiality terms that are consistent with the applicable protocol and any legal or safeguarding duties. Participants should be told whether statements may be attributed or quoted, how notes will be used and retained, and any limits to confidentiality. Clear terms support candid exchange while preserving the integrity and neutrality of the monitoring process.

Consultations differ in function and nature from hearings, observations and reviews of prosecutorial or judicial acts. The latter are directly observed or independently reviewed under a standard protocol, but they still require selection, interpretation and quality assurance. Consultations capture professional perspectives, reported experience and context. They may serve as evidence of those perspectives, but they do not, by themselves, prove that a procedural event occurred or establish the cause of an observed practice. Their value depends on source transparency, specificity, diversity and corroboration.

Consultations may strengthen analysis when they provide independent, relevant and sufficiently specific information that converges with other evidence. However, repetition is not automatically triangulation. Several actors from the same institution may repeat the same institutional account. A widely shared perception may still be incomplete or incorrect.

The practitioner should therefore record:

- 1) who was consulted by professional category;
- 2) the institutional perspective represented;
- 3) the questions or themes covered;
- 4) whether the account was first-hand or indirect;
- 5) areas of convergence;
- 6) divergent or minority views;
- 7) possible interests or limitations affecting the account;
- 8) which claims require corroboration through other sources.

Disagreement may be as informative as consensus, as it may reveal that responsibilities, rules or problems are understood differently across the system.



Method Note

In Collect, consultations gather professional perspectives, identify practical constraints, and highlight questions that other sources may test. Later sessions in Explain or Recommend may use consultations to present preliminary findings or recommendations for factual verification, challenge or co-design. These functions should remain distinct. Any preliminary material shown should be documented, and contrary or minority views should be preserved rather than treated as a failure to reach consensus.

c) Media monitoring

Media monitoring is a distinct yet complementary source stream that serves different functions at different stages. In Collect, it helps identify potentially relevant cases, preserves public statements, and documents the information environment surrounding monitored proceedings. It does not replace hearing observation, prosecutorial or judicial acts, or official records. Analyse may later assess patterns of prejudicial coverage or other fairness-related material. Explain may assess the media environment as a possible driver, pressure, or context for perception.

Practitioners should record and classify relevant media material linked to monitored cases and, where useful, broader reporting on the relevant case type and on justice-sector performance.

The media source pool should be intentional rather than accidental and should follow a specific collection plan. When applying this plan, practitioners may use:

- targeted monitoring of selected outlets;
- search alerts;
- media-monitoring platforms;
- case-specific searches;
- specialist investigative journalism sources;
- regional and local outlets;
- public statements and official social-media accounts, where relevant.

Media material may include:

- 1) articles directly connected to monitored cases;
- 2) publication of leaked, confidential or classified material;
- 3) articles relating to judicial actors;
- 4) investigative reporting relevant to the allegations or institutional response.

The sources should reflect the reform question and the information environment. They should not consist solely of the most visible national outlets, where local, specialist or investigative reporting is more likely to capture the relevant case.

A source log should capture at least the following: date, outlet, author where available, case link, procedural stage, type of claim, actors quoted, whether the article contains prejudicial language, whether it refers to confidential material, whether it reports statements by public officials or judicial actors, and whether the article reflects broader public-pressure dynamics.

The focus should be on content that:

1. May undermine the presumption of innocence of the accused; the reputation and/or security of victims, witnesses, attorneys and members of the judiciary involved in the processing of monitored cases.
 2. Use confidential sources, testimonies, classified evidence from monitored cases;
 3. Reports the views of influential persons expressing prejudicial views on the merit of a case (for example, a prominent politician accusing the judicial system of political bias in the processing of a monitored case).
 4. Reports the views of judicial actors on specific OCC cases.
 5. Instances where media are silent or clearly under-reporting on high level OCC cases (monitored or not, including at the investigation stage) when this situation can be reasonably traced back to possible or alleged political pressure and/or self-censorship
- As mentioned above, the purpose of the media analysis includes the contextualization of findings and the documentation of public perception of judicial response to a case. Relevant media coverage for these purposes may include reports not directly related to a monitored case, specifically:
6. OCC cases unveiled or investigated by the media, especially when publicly available evidence is not properly followed-up by the police and the prosecution;
 7. Stories of political pressure on members of the judiciary or law enforcement agencies;
 8. Instances of unethical behaviour by members of the judiciary or law enforcement agencies, including nepotism, cronyism.

SOURCE



Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

Media is a useful source of contextual evidence, including information on potential or pending cases, reported narratives or perceptions, pressures, and public statements. Do not use media as the source of procedural facts when primary or authoritative records are available. Media coverage may directly evidence what an outlet or public actor said, but it is not a proxy for population-level public opinion or confidence; those claims require appropriate perception data.

For example, an article stating that a hearing was postponed may direct the practitioner to the case record. It should not replace the hearing schedule, legal act, or direct observation when those sources are available. Likewise, practitioners may document prejudicial reporting directly as a media practice. A claim that reporting influenced a judicial decision requires separate evidence and should not be inferred merely from what may appear to be public pressure.

Media silence also warrants caution. Limited coverage may reflect editorial priorities, resource constraints, audience interest, or lack of access. Without further evidence, it should not be taken as proof of pressure or self-censorship.

5.2 Collection discipline: quality, provenance and evidence limits

a) Preserve source provenance and evidentiary limits

Assemble these materials, but do not collapse them into a single undifferentiated evidence pool. Direct observation, a prosecutorial or judicial act, an administrative statistic, a consultation account, media material and an evidence gap may relate to the same issue, but they do not necessarily establish the same proposition without further evidence. Consolidation must therefore preserve the function of each source rather than treating the number of sources as a substitute for evidentiary strength.

01 <i>Layer</i> Directly observed <i>What it carries forward</i> Hearing records from the Daily Hearing tool – what a practitioner saw, in person, at a public hearing.	02 <i>Layer</i> Written record <i>What it carries forward</i> Judicial acts, reviewed and coded against the reasoning and charging review criteria, including clarity, legal sufficiency and fact-element coherence.	03 <i>Layer</i> Case-connected <i>What it carries forward</i> The Case Report, linking hearings, acts and procedural developments across the life of each case.
04 <i>Layer</i> Statistical / distributional <i>What it carries forward</i> Official statistics, where available, tagged as distribution evidence rather than explanation.	05 <i>Layer</i> Perspective / contextual <i>What it carries forward</i> Consultation records and the media log, treated as complementary — never a self-standing basis for claims about institutional or courtroom practice.	06 <i>Layer</i> Absent or unverified <i>What it carries forward</i> The evidence-gap register, distinguishing impediments in access or collection, from the results of institutions' way of working that may have analytical value, such as what institutions generate, retain or make available.

b) Apply the tools consistently

When using the respective tools, practitioners should apply templates, tools and other materials to ensure consistency. This also requires training for those applying the rules, and the tools should be reviewed periodically. If a recurring inconsistency appears, the practitioner should correct the affected records, document the correction, and update the tools prospectively.

c) Assess evidence gaps

In institutions under pressure, missing or unavailable material can itself be informative. Missing material should be recorded as a fact, not silently treated as either an inconvenience or proof of institutional failure. Some gaps reflect temporary access or collection conditions, such as a file that is in use, not yet available, or accessible only through a route that has not yet opened; proportionate follow-up may close them.

By contrast, a gap in what the institution itself generates, retains or provides, where the information is required by law, procedure or a defined management function, may itself become part of the reform analysis. The practitioner should first describe the observable condition - non-production, non-access, delay, restriction, incompatibility or uncertainty - and reserve the question of whether it reflects capacity, access practice, reluctance or another condition for explanation. The register should record what could not be obtained, why, and whether the gap concerns project access or the institution's own information environment.

The register disciplines inference by preserving the distinction between “not observed” and “did not occur”. It also records when the gap prevents a claim, requires another source, triggers escalation, or justifies reopening Scope or Collect. A gap becomes a potential broader institutional or systemic analytical lead only after examining the relevant legal duty, institutional process, requests made, source fit and alternative explanations.

6 Assemble an evidence base

Collect closes by consolidating the separate source streams into a single, structured evidence base and transferring that evidence to Analyse. The product is a structured, source-tagged and modality-labelled evidence base, so Analyse begins with a single, legible body of evidence rather than five separate logs.

6.1 Convert collection prompts into analytical leads

The Case Report includes guiding questions on capacity, fairness and efficiency. During Collect, these questions focus attention and help the practitioner maintain issue lines. At hand-off, those issue lines should be presented to Analyse as analytical leads rather than as predetermined findings.

For **example**:

Analytical lead – possible capacity issue: Across three hearings, the prosecution struggled to link the financial expert evidence to specific charges. The judgement should be reviewed to determine whether and how the court addressed this relationship. The record does not yet establish whether the issue stemmed from prosecutorial preparation, expert methodology or the structure of the indictment.

This is sufficiently structured to guide analysis, but it does not pre-empt the analytical judgement.

6.2 Hand off an evidence base, not a conclusion

Analyse should receive a structured evidence base, which serves as the analytical foundation, containing both the evidence and the questions that remain open, including:

- the evidence strategy from Scope, so that the practitioner can see what the collection process was designed to establish;
- the confirmed case pool, including selection basis, observation window and monitoring modality;
- the source records, including Daily Hearing tools, Case Reports and legal-act reviews;
- the complementary evidence streams, including statistics, consultations and media logs;

- the evidence-gap register;
- the collection-completeness record;
- the provisional analytical leads, including possible concerns, good practices, contradictions and unresolved questions.

This hand-off should not direct Analyse towards a preferred conclusion. Its function is to make the evidence accessible, intelligible and contestable.

A well-assembled evidence base should allow the practitioner to move:

- from a proposed finding back to the relevant cases;
- from a case-level issue back to the specific hearings and acts;
- from a coded variable back to the document from which it was derived;
- from a professional explanation back to the consultation record;
- from an apparent absence back to the evidence-gap log.

That backward trace is essential to quality assurance.

► Before you move on



ETHICAL FRAME NOTE

Collect is the only Step conducted inside live proceedings, which underscores the importance of adherence to ethical principles.

Agreement. Access to schedules, records and counterparts requires agreement with authorities and should be distinguished from the public's legal right to attend open proceedings, where provided by domestic law. An access arrangement must not create an institutional veto over case selection, findings or lawful public observation. In this Step, agreement means that access is structured, transparent and institutionally authorized.

Non-intervention. At a minimum, it means that practitioners do not engage with the court on the merits of an individual case and do not attempt to influence the outcome of proceedings, directly or indirectly. If an interlocutor shifts the discussion towards the merits of a case, the practitioner should clarify that it cannot comment and redirect the conversation to the broader institutional issue. Non-participant observation reduces interference but does not eliminate observer effects or interpretive bias; practitioners should use unobtrusive conduct, use standard forms, and conduct periodic dual-observer calibration.

Confidentiality. Confidentiality safeguards the integrity of the evidence base and the safety and trust of those who engage with the monitoring process. It prevents the release of information gathered during monitoring, including sensitive material disclosed at a hearing, and private views expressed by parties or legal actors. Public-source status does not remove these duties. It also prevents ad hoc disclosure of information before it has been reviewed, analysed, and cleared through the agreed reporting channels. Confidentiality also means that practitioners do not seek or knowingly receive confidential documents. If confidential information is inadvertently communicated, existing protocols should be followed to protect, remove or destroy it, and it should not be entered into internal databases, records or reports.



COMMUNICATION CARD

At this Step communication is primarily about obtaining and clarifying data— not to secure endorsement of what that evidence may eventually show. What you can share is what you are examining, for what period and procedure, and which material would help. You are not obliged to answer "what are you finding?" before Analyse has done that work — the material is still being consolidated, and saying so is a complete answer. What you are getting back is information about how things are done in practice and where information exists. However, an institution's account of a practice is evidence of the account, not of the practice. Missing material, selective access or an unexpected source is recorded under Collect; if it changes the scope of the inquiry itself, return to Scope.



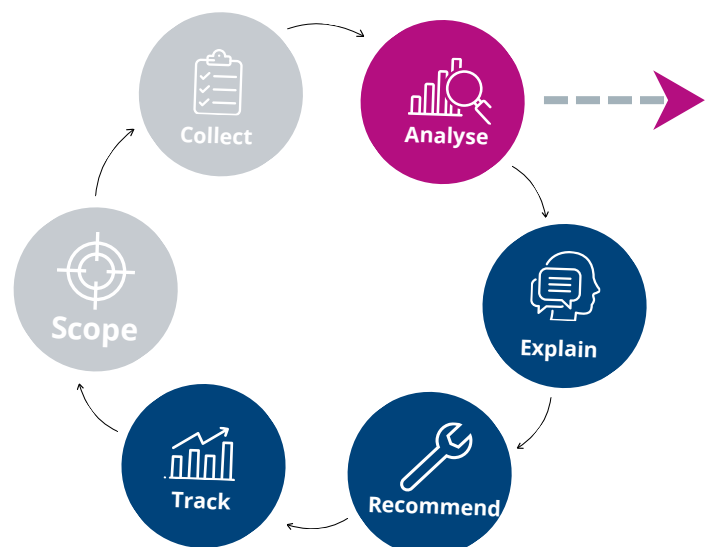
ADAPTATION NOTE

The collection discipline extends beyond criminal trials, including those related to serious organized crime and corruption. The distinction between core and complementary sources, provenance rules, evidence-gap discipline and ethical safeguards should be maintained. The adapted evidence strategy should explain which sources support direct findings, which support only context or reported experience, and which fall outside observation.



STEP-CHECKLIST

- ☐ Are **direct observation, judicial acts, statistics, consultations and media** analytically separated?
- ☐ Are the **Case Report issue lines linked to specific sources**?
- ☐ Are **contrary evidence and unresolved questions** preserved?
- ☐ Are **missed hearings, unavailable acts and unreliable statistics** recorded?
- ☐ Is it clear **what the evidence base cannot support**?
- ☐ Can **another practitioner reconstruct** how the evidence was collected and organized?



Step 3



Analyse

► Where you are in the cycle

Collect structures the source material, but at that point it has not yet been assessed as evidence for a specific claim. Analyse reads that material, first within individual hearings and cases, then across cases and themes, to produce bounded findings and cross-case patterns with the rigour necessary to support Explain and, ultimately, reform design. Analyse does not yet explain contributing conditions or design recommendations. Rather, it serves as a critical intermediary Step to determine what the material supports, for example, which concerns recur, which good practices appear, which patterns are visible, which claims can be made, and where the evidence remains too limited for a conclusion.



► Why this Step is important

Collected material does not speak for itself, and it certainly does not speak in the language of reform. Analysis is therefore the Step that turns collected material into a structured account of practice.

► What happens if we are not intentional about it

Without a disciplined reading, isolated occurrences are mistaken for systemic failures and vice versa. Not every procedural event is a pattern. Not every pattern is a cause. Not every concern is systemic. The result is predictably that the analysis is unreliable. Weak analysis yields conclusions that appear selective or impressionistic and cannot support credible recommendations.

► What you produce

By the end of this Step, the practitioner should have:

hearing-level, case-level and cross-case analytical findings and, where the threshold is met, systemic findings; concerns and good practices read through the dimensions of capacity, fairness and efficiency; a source map showing which evidence supports each finding; bounded findings that separate observation from inference; preliminary hypotheses about possible drivers; evidence gaps requiring further collection or a Deep Dive; and findings ready to anchor the explanatory inquiry in Explain.



THE STEP AT A GLANCE

1

Set the level of claim

What can this material speak to?

Place the level at which the evidence supports drawing a conclusion provisionally at one of four levels – single hearing or legal act, individual case trajectory, cross-case pattern, systemic finding – knowing that movement between them is never automatic.

2

Apply the dimensions of effectiveness

What does the evidence show against the effectiveness standard?

Read the material through Fairness, Efficiency, and Capacity, recording concerns and good practices alike.

3

Bound the finding

What can we responsibly say, and how strongly?

Hold to bounded inference: comparability, stated denominators, triangulation, contrary evidence, the analytical safeguards, and consistent confidence language.

4

Structure the finding

How is the claim written so it withstands scrutiny?

State the issue or practice, identify the applicable standard, test the evidence against it, and state the bounded conclusion with its scope and limits.

5

Route the finding

What moves forward, and what needs more?

Send the finding to Explain with its scope statement and source map, or hold on to it for further targeted collection or a Deep Dive.

► What analysis does: from source material to bounded findings

Analyse is the disciplined assessment of collected source material to produce bounded findings. It asks what the record can responsibly and rigorously support before the practitioner moves forward. This Step performs two related but distinct functions: First, it applies the three dimensions of effectiveness – Capacity, Fairness and Efficiency – to hearing observations, case reports, legal acts, statistics, consultations and other collected material. Second, it disciplines the resulting claims, ensuring that every finding is traceable, appropriately scoped and defensible.

Throughout this Step:

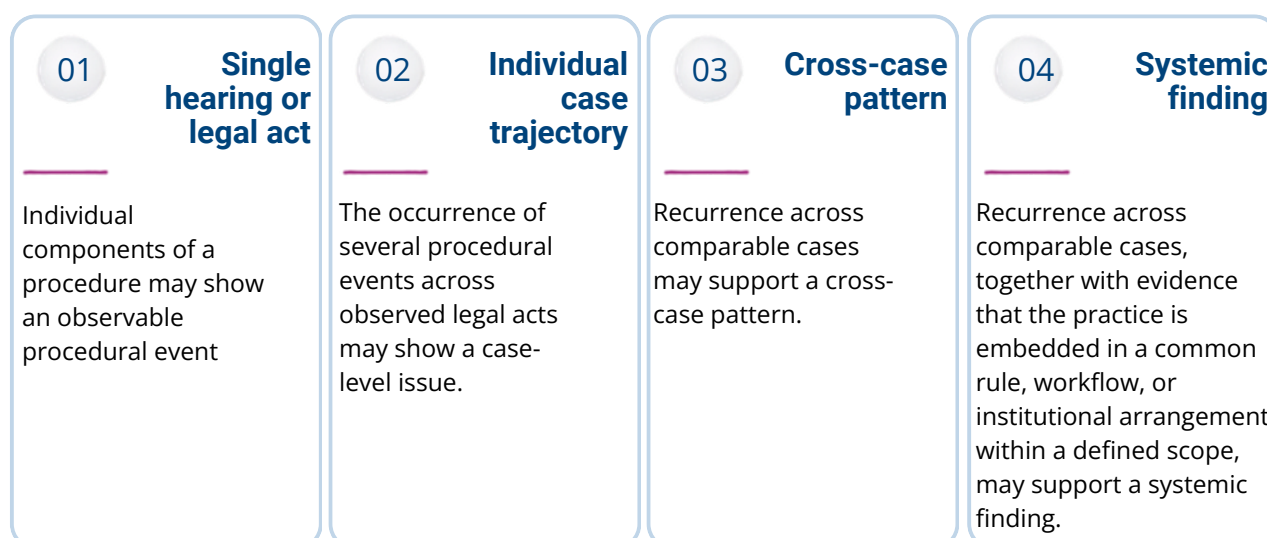
- Procedural events are specific procedural actions or decisions of potential relevance observable to the practitioner through hearings and legal acts;
- Source material refers to collected observations, records and coded variables;
- Evidence means source material assessed as relevant and sufficiently reliable for a specific claim;
- Findings are the bounded conclusions drawn from that evidence against a stated standard.

Analyse determines what the material supports, at what level, and subject to which limitations, based on observable practice. In doing so, it assesses observable practice against stated standards.

To do that, the practitioner should not start by asking what an issue “really means”, who is responsible for it, or what reform should follow. Those questions come later. The first task is more constrained: to determine what the source material can support, at what level, and with what limits.

1 Set the level of the claim

Before applying the effectiveness dimensions, set the provisional level of the claim. The same observation has different significance depending on whether it appears in a single hearing, within a case, or across comparable cases. There are four levels:



The movement between levels is not automatic; each step requires additional, relevant evidence.

a) Hearing level

At the hearing level, analysis identifies what happened during a specific procedural event: whether parties were present, whether the hearing was productive, whether a procedural event arose, whether a motion was made, whether the court ruled, whether the hearing advanced the case, and whether any concern or good practice was visible.

The task is to record the event accurately, identify its immediate procedural significance, and preserve the source.

How to analyse at hearing level

→**First**, record the event as precisely as possible, focusing on observation and established facts and avoiding inference. For example: the expert report was not ready; the court adjourned the hearing; the record attributed the adjournment to non-submission of the report; and the court scheduled the next hearing three months later.

→**Second**, identify the event's procedural significance. Did the hearing advance the case by producing a ruling, taking evidence, hearing a witness, addressing a motion, clarifying the indictment or preparing the next stage? Or did it repeat an earlier blockage?

→**Third**, classify the event in a way that can later inform the next levels of analysis. This is where qualitative and quantitative methods meet. Qualitative analysis is used to interpret the procedural dynamics and contextual factors that become visible during a hearing and may be relevant to the three dimensions of effectiveness: Capacity, Fairness and Efficiency. In parallel, selected features of the hearing are recorded against predefined variables and categories so that they can be compared and aggregated across hearings and cases. Where structured coding is used, apply the codebook or coding protocol established in Collect; Analyse should not invent new categories after seeing the results without documenting and justifying the change. For example, recording both the circumstances of an adjournment and its stated reason preserves the procedural context of the event. Classifying that reason consistently also makes it possible to examine the frequency and distribution of different sources of delay across the wider case pool. This is a distinctive contribution of EBTM: it preserves the detail of courtroom practice while structuring selected features of that practice for analysis across a broader evidence base.

The hearing-level finding should remain narrow unless other evidence allows it to expand.

For **example**:

The hearing did not advance because the expert report was not ready. The record attributes the adjournment to expert-evidence preparation. This is recorded as a hearing-level observation with potential efficiency significance for follow-up across the case trajectory.

This observation does not establish systemic delay. A broader cross-case or systemic finding requires further analysis at the appropriate level; weak court control or poor expert management are possible explanatory hypotheses only after a broader finding is supported. This formulation keeps the claim at hearing level and avoids extrapolation beyond what can be established.

b) Case level

At the case level, Analyse connects events over time and asks whether, taken together, hearing observations, procedural chronology and legal acts support a case-level concern or good practice.

The Case Report tool, described in Step 2, therefore does more than collect data; it also prevents practitioners from treating isolated hearing notes as final findings. Practitioners should update issues throughout the proceeding and link each point to the source. Step 2 sets out the distinct functions and limits of these source streams.

How to analyse at case level

→**First**, build the case chronology, including key hearings, adjournments, motions, rulings, evidentiary Steps, changes to charges, plea agreements, judgements, appeal decisions, and, where relevant, enforcement or confiscation developments.

→**Second**, track whether the same issue recurs. One adjournment recorded as linked to a late expert report may remain an isolated event; several adjournments in the same proceeding linked to expert evidence may support a case-level issue.

→**Third**, connect hearings and legal acts. The Daily Hearing record captures what happened in court, while a legal act may show how an issue was formally framed, reasoned, resolved or carried forward. Use each source only for the claim it supports.

→**Fourth**, state the finding at the case level. One case may reveal an important issue, but it does not establish a cross-case pattern on its own.

For **example**, a case-level finding may read:

Throughout the proceeding, repeated adjournments linked to expert evidence affected trial continuity and contributed to the case's duration. The case material may support a case-level efficiency concern and raise questions about expert-evidence planning but does not show whether the issue recurs across comparable cases.

This is broader than the hearing-level observation but remains properly bounded to one case.

-End-of-Case Assessment

A key component of case-level analysis is the End-of-Case Assessment. While hearing and case reports follow the proceeding as it moves forward, the End-of-Case Assessment reviews the pathway retrospectively from the closure point defined by the protocol. The practitioner must explicitly record the procedural stage and outcome status, including whether it is first-instance, final, final and binding, discontinued, or otherwise closed.

The End-of-Case Assessment begins by recording the procedural result – conviction, acquittal, partial conviction, sentence, confiscation, appeal outcome, retrial, discontinuance or another form of closure – and then examines what the record shows about the pathway to that result: indictment structure, fact-element mapping, legal qualification, evidentiary issues, financial evidence, hearing management, reasoning and appellate findings.

The purpose is not to make one case prove the system. It is to understand the case-level trajectory.

The practitioner asks:

What does the record show about the procedural, evidentiary and reasoning pathway to this outcome, and what remains uncertain?

This retrospective reading is valuable because some issues become clearer only at the end of the case. An issue of charge accuracy may become more apparent in the judgement, especially if the court identifies the defect. A delay in the financial investigation may help explain the absence of confiscation only where the record supports that link. Issues with the soundness of legal reasoning may become clear when an appellate court upholds, reverses or criticizes the first-instance decision.

The End-of-Case Assessment must guard against outcome bias. A result should not cause the practitioner to read the proceeding backwards as if conviction, acquittal, confiscation, sentence or appeal outcome were self-explanatory. The assessment must not treat conviction as success, acquittal as failure, or confiscation as an automatic marker of effectiveness. It should record what the court held, which evidence it assessed, how it reasoned, which safeguards were observed and which legal tools were used, without substituting the practitioner's own view of the merits.

A two-pass review helps avoid hindsight risk. First, assess the indictment, hearings and judicial acts against predefined criteria, without using the outcome as the organizing hypothesis. Second, record the result and examine the pathway retrospectively, noting any divergence between the two assessments. Distinguish what the record shows, what it suggests, which rival explanations remain, and what cannot be determined. Where the pathway is unclear, contested or dependent on factors outside the monitored record, the assessment should not force a conclusion. It should record the uncertainty, triangulate with other sources where possible, and, where the issue recurs across cases, consider whether it should trigger a Deep Dive.

The End-of-Case tool produces a case-level trajectory assessment. When aggregated or analysed separately across cases, these assessments may support the identification of cross-case patterns, higher-resolution Deep Dives, and preliminary diagnostic hypotheses for Explain; the tool does not perform those later analytical functions on its own.

c) Cross-case level: analyse recurrence across comparable cases

At the cross-case level, analysis asks whether case-level findings recur across a wider pool. This is where the practitioner begins to identify patterns, common bottlenecks, repeated vulnerabilities, divergent practices or replicable good practices worth preserving. A cross-case pattern requires more than multiple examples; it requires a defined and defensible comparison.

How to analyse at cross-case level

→**First**, define the comparison group. Are the cases comparable by offence type, defendant status, procedural route, court level, case complexity, legal tool, procedural phase or outcome? A pattern in contested high-status corruption trials should not be casually merged with a pattern in low-level plea-agreement cases.

→**Second**, identify the repeated issue precisely. “Inefficiency” is too broad. “Repeated adjournments linked to delayed expert evidence in multi-defendant corruption trials” is closer. “Weak capacity” is too broad. “Repeated failure to connect facts, legal elements and financial evidence in indictments and judgements” is more precise.

→**Third**, disaggregate the pattern. Ask where the issue gathers, for example:

- by procedural phase;
- by actor or institution;
- by court or chamber;
- by case type;
- by defendant status;
- by charge gravity;
- by procedural route;
- by outcome.

→**Fourth**, state the pattern without explaining it yet. Cross-case analysis can show recurrence. It can generate hypotheses. It should not yet assert the underlying driver or root cause of the identified pattern.

Every cross-case statement should define the eligible comparison group, identify any missing or excluded observations, and state whether multiple defendants or charges are nested within the same case. ‘Several cases’ is not an adequate description of coverage.



RTM Application

In RTM, analytical cross-case leads are carried into a separate excel-based “issues log”. The log is structured using an adapted IRAC scaffold (framework used for legal analysis: Issue, Rule, Application, and Conclusion) and provides the working space in which the practitioner tests the issue against the applicable standard, assembles supporting and contrary evidence, bounds the inference and develops the eventual finding and later on links it to a policy reform recommendation.

A cross-case finding may read:

Across [n] of [N] eligible monitored multi-defendant corruption cases, adjournments linked to expert evidence recurred at trial stage and contributed to longer proceedings. The pattern was most visible in cases involving financial evidence. The evidence supports a cross-case efficiency concern and raises questions for Explain about expert availability, court control, party preparation and institutional co-ordination.

End-of-Case Assessments can also support cross-case comparisons. They may, for example, show whether final decisions or appellate findings repeatedly identify gaps in fact-element mapping, whether confiscation is repeatedly absent after conviction, or whether appellate review repeatedly identifies the same reasoning weakness.

A recurring but under-resolved issue does not automatically warrant a Deep Dive. Many patterns should remain as bounded findings or evidence gaps. A Deep Dive may be appropriate only where the question is material to a monitoring or reform decision, additional inquiry has decision value, the method is tractable, the cost and risk are proportionate, access is feasible, and the inquiry is authorized.

d) Systemic level: decide whether the pattern reveals a routine institutional practice

A systemic finding goes beyond recurrence across cases. It requires evidence that the recurring practice is embedded in, governed by, or traceable to a common rule, workflow, routine, or a formal or informal working arrangement within a defined institutional or procedural scope. Establishing the descriptive institutional reach remains an Analyse task; Explain then tests why the practice occurs or persists and which contributing conditions support it. A formally adopted or informally practiced institution-wide procedure, verified in operation, may support a systemic finding even where observation points are limited, provided the finding states the evidentiary basis and scope.

How to analyse at the systemic level

The practitioner should avoid asking only how many cases showed a concern. Four checks decide whether a pattern earns the label “systemic”.

→**First**, test for recurrence. Does the issue appear across enough comparable material to move beyond an isolated case or anecdote?

→**Second**, define the pattern’s institutional reach and scope. Does it reflect a common rule, workflow or practice within a specialized court, a procedural stage, several offices, or a wider jurisdiction? Recurrence across multiple actors is relevant but not mandatory; repeated observations involving one actor may still support a systemic finding where independent evidence verifies that the issue is embedded in a common rule, workflow or institutional arrangement within the defined scope; the scope of generalization must not extend beyond that evidence.

→**Third**, test whether the evidentiary base is sufficiently corroborated for the breadth and nature of the systemic claim. Broader institution- or jurisdiction-level claims will ordinarily require a genuinely independent or meaningfully complementary evidence stream, such as official statistics, institutional documents, judicial-act review across a wider pool, or another structured source. The question is not whether an arbitrary second source exists, but whether the evidence establishes institutional reach rather than repetition within the same underlying record. Media content and consultations may directly evidence published statements, reported perspectives or contextual conditions and may corroborate other claims where source fit permits; they will not ordinarily, on their own, establish a systemic claim about courtroom or institutional practice. Tier 2 document-based monitoring can strengthen coverage, but it is not independent corroboration where it draws on the same underlying record.

→**Fourth**, test whether the pattern is confined to a subgroup or remains visible across the specified disaggregation. This may use quantitative tables or a structured qualitative comparison. Disaggregation tests scope and robustness; it is not synonymous with quantitative analysis and does not establish causation. Quantitative analysis also requires clear denominators, rules for missing data, and attention to observations nested within defendants and cases.

Where disaggregation is warranted, the practitioner should select the axes best suited to testing the scope and boundaries of the emerging pattern. The table below illustrates common dimensions of comparison and the analytical question each can help answer.

01 *Disaggregation axis***Procedural phase***Questions to ask*

Does the concern appear at indictment, confirmation, trial, sentencing, appeal, retrial or enforcement?

Why it matters

Locates the decision point where the pattern emerges.

02 *Disaggregation axis***Case type and charge***Questions to ask*

Does it appear in corruption, organized crime, money laundering, confiscation-related or mixed cases?

Why it matters

Prevents overgeneralizing from one case type.

03 *Disaggregation axis***Defendant/charge unit***Questions to ask*

Does it affect high-status defendants, high-gravity charges, peripheral defendants, or specific charges differently?

Why it matters

Avoids flattening mixed outcomes in multi-defendant cases.

04 *Disaggregation axis***Procedural route***Questions to ask*

Does it appear in contested trials, plea agreements, retrials or appeals?

Why it matters

Shows whether the route shapes the pattern.

05 *Disaggregation axis***Time period or legal framework***Questions to ask*

Does the pattern persist across relevant periods, procedural regimes or policy changes?

Why it matters

Defines the temporal scope and may reveal context-specific variation.

06 *Disaggregation axis***Institutional setting***Questions to ask*

Does it appear in one court, several courts, one jurisdiction, or regionally?

Why it matters

Does it appear in one court, several courts, one jurisdiction, or regionally?

This analytical process has three possible outcomes. The pattern may be confined to a subgroup or setting, such as a single case type or a single court. It may remain robust across the tested disaggregations. Or it may be indeterminate at the current unit or resolution because the available data cannot reveal the distinction that would resolve the question, for example, when case-level outcomes obscure charge-level variation. An indeterminate result may be routed to the Deep Dive if the trigger criteria are met.

A systemic finding may read:

Across the monitored pool of complex corruption cases, repeated difficulties in integrating expert financial evidence arise at the trial stage and are associated with adjournments, evidentiary disputes and limited reasoning on illicit gain. Independent or meaningfully complementary evidence, such as [specified administrative data, institutional workflow evidence or wider legal-act review], also shows that the relevant practice or arrangement operates across [defined court(s), institution(s) or procedural scope]. Taken together, the evidence supports a systemic finding within that stated scope, specifically at the stage where expert financial evidence is introduced and assessed.

Because the pool is purposely selected, this finding does not estimate prevalence across all corruption cases and the finding does not establish whether the driver lies in prosecution preparation, expert availability, court control, workload, legal design or institutional co-ordination.

The distinction can be summarized as a progression in analytical reach: each level asks a different question, draws on a broader evidentiary base and supports a correspondingly broader claim. The table below shows how those levels relate without implying that every issue must travel all the way to a systemic finding.

In a snapshot

 Level	 Analytical question	 Typical sources	 Output
Hearing level	What happened during this procedural event?	Daily Hearing record, courtroom observation, minutes or transcript where used.	Event observation; procedural event; possible concern or good practice.
Case level	Do repeated events, judicial acts and the procedural chronology show a case-level issue or good practice?	Case report, hearing summaries, procedural chronology, legal acts, media/context material.	Case-level analytical finding, always traced back to source.
Cross-case level	Does the case-level finding recur across a wider pool?	Case reports; judicial-act reviews; monitoring dataset; statistics; cross-case coding; End-of-Case Assessments; deep-dive datasets where available.	Cross-case pattern with the comparison group, coverage and limitations stated.
Systemic level	Within what defined institutional or procedural scope does the cross-case pattern recur, and what evidence shows that it reflects a common rule, workflow, routine or institutional practice within that scope?	Cross-case findings; disaggregated statistics; repeated End-of-Case trajectories; legal-act patterns; institutional documents; appropriately designed consultations; reform-tracking evidence; Deep Dive analysis.	Systemic finding with an explicit scope, evidentiary base and limitations; possible explanatory hypotheses routed to Explain.

The table should be read from top to bottom. A finding may move from one level to the next only when additional evidence allows it to travel. The practitioner should therefore ask, at each level, what new evidence justifies raising the claim's level. If the answer is unclear, the claim should remain at the lower level or be recorded as an evidence gap.

1.1 Sources and their evidentiary value

a) The two primary source types

i. Written acts

As reflected in the ODIHR-developed trial-monitoring methodology for the RTM Project, written acts are a central source for assessing capacity of judges and prosecutors. These include indictments, first instance verdicts and sentencing, and decisions on appeal (verdict and sentencing), which can be viewed as key decision points in any type of criminal proceeding.



The evaluation of the quality of judicial decisions can and should be kept distinct and separated from any process aimed at checking and second-guessing the reconstruction of facts and interpretation of the law by judges and prosecutors; this is a scope of review which can be properly conducted by an appeal court; its adoption in the context of trial monitoring, however, would overstep the limits of appropriate oversight over the work of the judiciary and arguably impinge upon its independence.

Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project, summarized.

Review should assess clarity, coherence, responsiveness to material submissions and the use of applicable law, without substituting the practitioner's view of the merits. Written acts cannot, on their own, show the full course of a proceeding or explain why a deficiency occurred. Their methodological advantage is replicability: multiple reviewers can apply the same criteria to the same text, including through blinded review where feasible.

Written acts are therefore important yet incomplete. They record what was decided and how the decision was formally reasoned; they may not show how the case was managed, how evidence was introduced, how witnesses were protected or examined, how the court responded to procedural disorder, or how the parties used available tools in practice. EBTM addresses this limitation by analysing written acts alongside in-court performance.

This dual-source basis matters because serious and complex cases develop through both written acts and in-court practice before the final decision. A judgement may reveal a thin evidentiary theory, while hearing records may show how that theory was presented and tested. The combined record may show that confiscation was not pursued, that early planning did not occur, or that witness-protection issues affected testimony. These observations may raise questions about capacity, fairness or efficiency. They do not, without further analysis, establish why the practice occurred.

ii. In-court performance

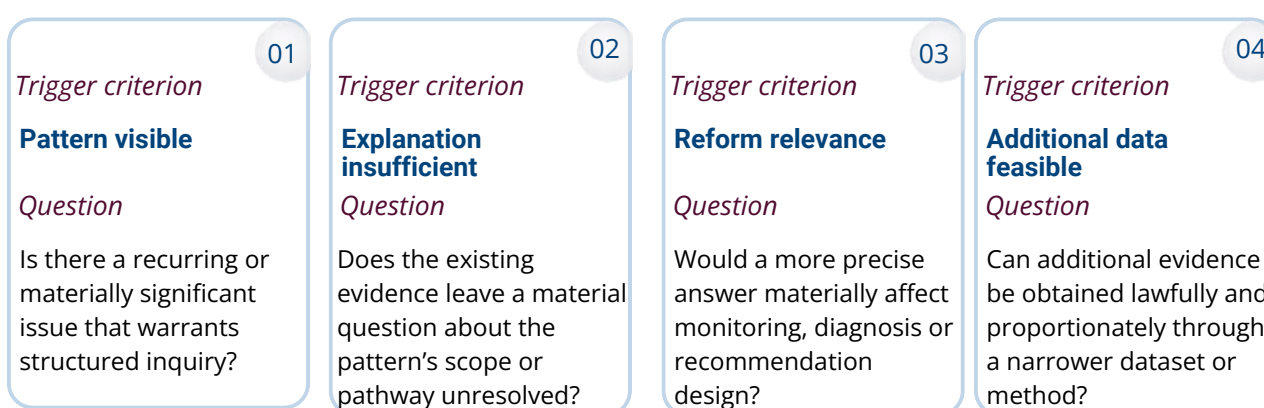
In-court performance, as captured by observation of hearings, complements written acts by showing how the proceeding was conducted: how the court managed hearings, how the parties presented and challenged material, and how procedural safeguards operated in practice. Observed conduct may raise questions about capacity, fairness or efficiency, but the underlying condition remains to be tested in Explain. Courtroom conduct may reflect legal design, workload, co-ordination, party strategy, professional practice or lawful discretion.

Both source types feed all three lenses. They indicate where relevant material may be found, not which lens applies. A single act or hearing may support more than one line of inquiry; the relevant dimension in §3.2 determines the question posed to it.

b) The wider evidence base

i. Deep Dives

A Deep Dive may be warranted when existing evidence reveals a material recurring issue but lacks the resolution needed to determine its scope or to support a responsible handover to Explain (more details on the methodology and function of deep dives can be found in Annex 1). The interpretive lens will usually already have exposed the question. A trigger may arise from hearing observations, legal acts, statistics, case reports, media context, consultations, or tracking data.



Examples include:

- conviction rates that hide charge-level or defendant-level attrition;
- plea agreements that resolve cases but leave material questions about the factual basis, sanctions, co-operation or confiscation;
- confiscation gaps that require tracing the financial-investigation chain; and
- transparency gaps that require comparing formal publication rules with actual access to schedules and decisions.

Many patterns should remain as bounded findings or logged evidence gaps. Where the issue is material, tractable, proportionate and authorized, route it to the Deep Dive to rescope, collect and analyse at higher resolution, then return the specialized findings to Explain alongside the trial-monitoring findings.

ii. Use of media

Media enters analysis in three distinct ways, which must not be collapsed: as a lead or contextual source requiring verification; as an object of analysis, where coverage or public statements may pose a legal or ethical risk; and as a window into how justice institutions communicate and function. Serious organized crime and corruption cases often attract competing narratives of accountability, impunity, political revenge, selective justice, institutional weakness, or elite protection. Those narratives show how specific outlets frame a case; they do not determine whether the proceeding is fair or establish population-level perceptions of independence or public confidence.

The governing limits are clear: public interest is not pressure; criticism is not interference; and media attention is not proof that a court or prosecution was influenced. Media content reflects media narratives, not public opinion, public confidence or perceived independence across the population; those claims require appropriate perception data.

Unless distinct and sufficient evidence in the monitoring record supports that causal claim, the practitioner should avoid asserting that the media influenced the court; such evidence will rarely be available.

A bounded formulation would be:

Media coverage of the case included repeated public statements by public officials calling for a conviction while the criminal proceedings remained pending. This raises a contextual fairness concern regarding the presumption of innocence and possible external pressure on adjudicative independence. The available evidence does not show whether the statements affected judicial decision-making.

Where media-related concerns recur across comparable cases, they may support a cross-case pattern or trigger a Deep Dive into transparency, institutional communication, or the public environment surrounding adjudication.

c) Source fit and probative value

Identify the source for every claim. Its value depends on the claim it supports in terms of its directness, reliability, independence, and fit, rather than on a fixed hierarchy. A hearing observation may support what occurred in court; a judgement may support what was formally reasoned; and a party submission may show what a party alleged, not that the allegation is true.

d) Verification and triangulation

Every analytical claim should identify its source and provenance. The source map should record the origin, modality, date, directness, quality, independence, and the precise claim supported. Sources are not independent merely because they appear in different formats; several reports may repeat the same underlying account. A direct hearing observation may support a claim about courtroom management. A judgement may support a claim about reasoning. A procedural chronology may support a claim about timing. A statistic may support a claim about distribution. A consultation may support a claim about professional perception or a possible explanation, but it should not stand alone as proof of a pattern of practice.

Verification asks whether a source supports the specific claim. Triangulation asks whether genuinely independent or complementary sources converge, contradict each other, or reveal different aspects of the same practice. Repeated accounts of the same underlying account are not triangulated. The table below provides useful practical insights using plausible findings.

 Finding	 Source 1	 Source 2	 Source 3	 Triangulation result
Repeated adjournments due to absent witnesses	Daily Hearing records show postponed sessions.	Procedural chronology shows repeated delay.	Judicial act or ruling records reason where available.	Sources align and may support a case-level or cross-case efficiency concern, depending on coverage; they do not explain the cause.
Thin confiscation reasoning	The judgement provides limited reasoning on illicit gain.	The indictment contains a limited or unclear financial theory.	The appeal decision identifies, confirms or does not address the same issue.	Sources may illuminate the pathway; the possible driver remains for Explain.
Possible public pressure	Media log shows prejudicial statements.	Hearing observation shows the courtroom environment.	Consultations report perceived pressure.	Contextual risk only; do not infer an effect on adjudication without distinct evidence.

Evidence may be corroborative when an independent source confirms the same fact; complementary when a different source adds a related fact; or redundant when it repeats what is already known without providing independent confirmation. Corroborative and complementary evidence may strengthen a claim. Redundant evidence should not be counted as if it did.

Record conflicts among sources rather than smoothing them over; the conflict may be analytically useful and ignoring it may undermine the validity of the observation and the findings. Where the evidence shows a pattern but does not support a broader or explanatory conclusion, retain the narrower finding and state the limitation.

2 Applying the dimensions of effectiveness

Once the practitioner has set the provisional claim level, the three dimensions provide the normative and performance framework for interpreting it. The claim level answers how far the evidence extends; the lens answers why the observed practice matters.

The three dimensions – Capacity, Fairness and Efficiency – are distinct yet overlapping aspects of effectiveness.⁶

⁶The RTM Project methodology treated strategic use of judicial tools as a separate fourth dimension for purposes of reporting. This Guide retains that content as a cross-lens strand running across Capacity, Fairness and Efficiency.

Together, they frame the central EBTM question: *does the justice system handle serious and complex cases lawfully, fairly, competently, and within a reasonable time, using its authority and resources appropriately?*

The sequence matters. Do not label an observation as a capacity problem, a fairness concern, or an efficiency issue before determining which level of claim the material supports. The lens can be applied at each level, but the claim level determines how far the interpretation may extend.

The lenses also prevent false proxies. Effectiveness cannot be reduced to conviction rates, speed, case volume or the mere existence of legal tools. A conviction does not, by itself, show that justice was effective; an acquittal does not, by itself, show failure. A fast trial may be unfair or poorly managed, while a long trial may be justified by complexity. A tool written into law is not the same as one understood, lawfully used and integrated into practice.

Capacity, Fairness and Efficiency therefore help the practitioner interpret evidence without flattening it. The same practice may matter in several ways. A weak indictment may raise a capacity concern where the applicable domestic requirements or defined review criteria call for linkage among the facts, legal elements and evidentiary basis and that linkage is unclear; it may raise an efficiency concern where the record links it to amendments, disputes, retrial or appeal; and it may engage fairness if the accused faces serious charges that are not clearly framed.

The dimensions should not be treated as inherently independent. Begin with the dimension most directly engaged or relevant, then consider the implications under the others. This is particularly important in serious organized crime and corruption cases, where complexity, fair-trial guarantees, professional practice and public accountability often converge at the same decision point.

The sections below briefly explain each dimension before setting out its practical application.

 Dimension	 Core question	 Evidence sources	 Risk if misread
Fairness	Does observed practice respect the applicable fair-trial guarantees and related safeguards? Treat victim and witness protection, participation and broader transparency under their distinct legal bases rather than as interchangeable fair-trial rights.	Hearing observations, prosecutorial and judicial acts, case trajectories and, where claim-fit, statistics, institutional documents and other verified sources.	Treating every procedural irregularity as a fair-trial violation without applying the right-specific test; conversely, using overall fairness to erase a breach of a guarantee that has its own legal test.
Efficiency	Are time, complexity and resources managed so proceedings progress within a reasonable time without compromising fairness, legality or decision quality?	Hearing observations, prosecutorial and judicial acts, case trajectories and, where claim-fit, statistics, institutional documents and other verified sources.	Equating speed with effectiveness; ignoring case complexity; counting hearings without asking whether they advanced the case.

Capacity

Does the monitored record show that relevant justice actors and institutions can perform the required functions lawfully and coherently, including the technical, organizational, co-ordination, resource and enabling conditions relevant to the function?

Hearing observations, prosecutorial and judicial acts, case trajectories and, where claim-fit, statistics, institutional documents and other verified sources.

Second guessing the decision-making process of judicial actors in specific cases.

2.1 Fairness - adherence to fair trial standards

Fairness concerns whether criminal proceedings are conducted in accordance with applicable fair-trial rights and procedural safeguards, ensuring that parties can participate effectively and that decisions are reached through a lawful and impartial process. Its core comprises the accused's right to a fair and public hearing before an independent and impartial tribunal, including equality of arms, effective participation, defence rights, the presumption of innocence, and reasoned decision-making where required. Victim and witness participation and protection also matter, but they stem from distinct legal bases and must be balanced against defence rights. Broader transparency and public-confidence questions may be relevant to reform but should not be treated as interchangeable with fair-trial guarantees. Fairness is assessed through the process, not inferred from conviction or acquittal.

"Fair trial standards as developed by the ECtHR include what is arguably the best set of criteria for assessing in a direct, fact-based manner the level of independence and impartiality of the judiciary. Specifically, they give us a framework to look for signs of bias in the actual behavior of a judge in court or for the existence of objective facts (for instance conflicts of interest) which raise doubts as to his impartiality. Fair trial standards can also effectively pinpoint flaws in the legal framework particularly in the criminal procedure. In addition to this, it can be argued that respect of the rights of the accused is a strong sign of the impartiality of the judiciary especially when individuals who are not connected to the government are under trial."

SOURCE



Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project.

In this Guide, the practitioner applies the applicable legal baseline to observable practice. The question is not whether a trial 'felt fair', but whether conduct, procedural decisions, judicial acts, appellate findings, or verified contextual evidence indicate a concern about the legal conditions under which the proceeding unfolded.

Fairness concerns may arise in courtroom conduct, written decisions, the procedural treatment of parties, the use of intrusive legal tools, the scope of appeal review or the public environment surrounding a high-interest proceeding. The Case Report and Daily Hearing tools therefore require practitioners to identify the applicable right or safeguard and cite the source for each issue raised.

Fairness analysis can be structured around inquiry areas as follows:



Inquiry area



Guiding questions

**Independence
and impartiality**

Did observable facts raise a reasonable question under the applicable subjective or objective impartiality standard? Consider conflicts of interest, prior involvement, ex parte contact, unexplained panel changes or a pattern of materially different treatment between parties. Robust case management, sceptical questioning, an adverse ruling or tone alone do not establish bias or pre-judgement. Was any disqualification motion filed, and how was it resolved?

**Equality of arms
and effective
legal assistance**

Did each party have a meaningful opportunity to present its case and respond to motions, evidence and proposals? Did the defence have adequate time and facilities, access to material as required by the applicable standard, and practical and effective legal assistance? Where representation changed or was interrupted, did that affect effective participation? Where additional time was requested, how did the court respond? Were restrictions on participation lawful, reasoned and proportionate?

**Reasoned
procedural
decisions**

Where the applicable law or relevant fair-trial standard required reasons, were reasons discernible from the hearing record or a subsequent written act and sufficient for the decision in question? Did the court address material submissions to the extent required by the applicable standard? Were comparable requests treated consistently, subject to relevant factual differences?

**Public hearing
and transparency**

Where the applicable legal standard required a public hearing at that procedural stage, was it public? Where the public or media were excluded, was the restriction grounded in a permissible legal basis and no broader than the applicable human-rights standard allowed? Were the reasons apparent or recorded to the extent required by the applicable standard? Separately, did the institution meet applicable rules on schedules, basic case information and publication of decisions, and how did any verified access gap affect scrutiny of the proceeding?

**Presumption of
innocence and
public pressure**

Did a public authority make a statement that expressed or implied the accused's guilt while the criminal proceedings remained pending, potentially engaging the presumption of innocence? The assessment should consider the wording, context and procedural stage. Did other public statements or campaigns call for a particular outcome in a manner that raised a separate concern about pressure or independence? Did coverage disclose protected information, attack witnesses or parties, or otherwise affect the adjudicative environment?



Inquiry area



Guiding questions

Protection of witnesses, victims and vulnerable participants

Where a party sought protective measures for a witness, victim or vulnerable participant, were the measures requested, granted or refused, and on what stated basis?

Did measures such as anonymity, screens, in-camera testimony, video-link or exclusion of the public go no further than necessary to protect the individual?

Did the record show how those measures were reconciled with the accused's right to examine witnesses and challenge the evidence against them?

Where anonymity, absence or another protective measure materially restricted confrontation, did the assessment consider the importance of the evidence and the counterbalancing safeguards required by the applicable fair-trial standard?

Where a vulnerable participant testified, was the process adapted to that person's needs, and was any resulting limitation on the defence proportionate and explained?

Use of intrusive or specialized legal tools

Where the proceeding relied on intrusive or specialized legal tools – such as special investigative measures, undercover techniques, communications interception, or protected or anonymous evidence – was the measure authorized by the competent authority under the applicable legal framework?

Which legal test applied to the measure? Where the applicable domestic or human-rights standard required necessity, subsidiarity or proportionality, did the record show that those requirements were addressed?

Where resulting material was used in the criminal proceedings, were the applicable fair-trial safeguards observed, including a sufficient opportunity to challenge the evidence or adequate counterbalancing safeguards where disclosure, confrontation or anonymity restrictions were lawfully imposed?

Were the authorization, supervision, time-limit, review or other safeguards required for that measure observably applied?

Appeal and remedy

Did appellate review identify fair-trial violations, reasoning gaps or procedural defects? Were remedies effective? Did repeated appellate findings suggest a broader fairness concern across cases?

How fairness findings should be stated:

A fairness finding should identify:

- the right, safeguard or procedural condition engaged;
- the source showing the concern;

- the level of claim;
- the recurrence and coverage that support the selected claim level;
- whether the evidence supports a finding or only raises a question;
- what remains uncertain.

For **example**:

In [case/hearing], the court rejected a defence request for additional preparation time following the introduction of new evidence. The applicable law required the court to give reasons, yet none were discernible from the hearing record or any subsequent written act. This raises a case-level fairness concern linked to equality of arms and effective participation in the defence. The available evidence does not support a conclusion about motive or the effect on the outcome.

Or, at cross-case level:

Across [n] of [N] eligible monitored contested corruption trials, defence requests for additional preparation time following late evidentiary submissions were rejected without any discernible reasons in the hearing record or in a subsequent written act. The evidence supports a cross-case fairness concern linked to equality of arms and reasoned procedural decision-making. It raises questions for Explain about scheduling pressure, procedural rules, judicial practice and party conduct.

2.2 Efficiency

Efficiency concerns whether the justice system manages time, complexity and resources so that proceedings progress within a reasonable time without compromising fairness or quality. It is not raw speed. Serious and complex proceedings may legitimately take longer, and delay must be assessed in context, including the case's complexity, the conduct of the authorities and parties, and what was at stake. At the system level, efficiency also concerns the productive use of hearings, sequencing, co-ordination and the avoidance of unnecessary rework.

The ODIHR-developed project trial-monitoring methodology defines efficiency as the ability of the justice system to actively manage procedural complexity and effectively allocate resources to ensure a trial within a reasonable time.

In fact, efficiency is a central element in most institutional evaluations of judicial performance. It is typically assessed through standardized indicators, such as disposition time, clearance rate, and the number of pending cases, often combined with benchmarks aimed at establishing optimal and predictable time frames for the resolution of cases. These indicators, as reflected for instance in the EU Justice Scoreboard,⁷ provide a useful overview of the functioning of judicial systems.

SOURCE



Source: ODIHR-developed TM methodology, specifically tailored for the RTM Project.

Such indicators are particularly useful for identifying patterns and potential bottlenecks, stages of proceedings where delays are more frequent, and trends across different categories of cases. Whether these rise to a systemic finding depends on the additional evidence of institutional reach.

⁷ https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/eu-justice-scoreboard_en

EBTM adds value by examining efficiency at the practice level. Hearing observations, case reports, procedural chronology, judicial acts and outcome analysis show not only how long a case took but also how time was used. Analyse identifies where delay, procedural friction or productive case management becomes visible; Explain tests why it occurs.

Efficiency analysis therefore begins with an assessment of hearing productivity and case movement. A hearing is not productive merely because it took place. It may advance the case by taking evidence, hearing a witness, resolving a motion, clarifying a procedural issue, ruling on admissibility, setting a credible timetable, or preparing the next stage. It may also add delay if it is adjourned without substantive activity, scheduled at long intervals, postponed for avoidable reasons, or undermined by lack of preparation.

Efficiency also has a quality dimension. High output or rapid movement may mask fragility, where recurring avoidable errors lead to remittal or retrial, appellate decisions repeatedly identify the same procedural defect, judgements lack the reasoning required by law, or financial aspects remain unresolved. Appeals, remittals and retrials are not, in themselves, inefficiency; they may be essential safeguards. The concern is repeated rework or delay attributable to identified and avoidable defects. Efficiency in serious and complex cases is therefore the disciplined management of proceedings to support lawful, fair and reasoned adjudication.

Where monitoring data permits, efficiency analysis should report stage-specific duration, intervals between hearings, hearing productivity, adjournment rate and coded reasons for adjournment, indictment-to-decision time, and appeal or retrial duration. For each indicator, state the denominator, coverage period, and missing-data rule. These measures describe timing and distribution; they do not establish cause. Efficiency analysis can be structured around inquiry areas as follows:

Inquiry area

Guiding questions

Overall duration and procedural stages

Was the duration of a monitored stage potentially excessive considering the applicable standard and the case circumstances? Did long periods pass between filing, any confirmation or review stage, plea, trial, appeal or retrial? Were statutory or procedural time limits exceeded, and are the reasons visible in the record? Findings about investigation duration require investigation-stage records or clearly attributed judicial or appellate findings; absence from the trial record is not proof of investigative delay or non-performance.

Hearing productivity

Did hearings advance the case through evidence-taking, witness examination, rulings, procedural clarification or concrete preparation for the next stage? Were many hearings held without substantial activity? If hearings were unproductive, what reason was recorded?

 Inquiry area

 Guiding questions

Adjournments and intervals

Were hearings frequently adjourned? Were they scheduled at long intervals? Was the trial adjourned to an indeterminate date? Were reasons recorded clearly? Were adjournments linked to witnesses, experts, parties, defence counsel, prosecution preparation, court scheduling, translation, prison transfer, illness, strike, security, or other factors?

Judicial case management

Did the judge or panel use lawful case-management measures to prevent avoidable delay while preserving defence rights? Do not classify the exercise of defence rights as a delaying tactic merely because it extends proceedings. Any adverse classification should rest on observable timing, repetition, legal relevance and, where appropriate, a reasoned judicial finding of abuse.

Planning in complex cases

In cases with multiple defendants, large volumes of material, expert or financial evidence, was a status conference or equivalent planning mechanism available and used where appropriate? If used, did the record show that it produced a realistic timetable, clarified sequencing or reduced disruption? If it was not used, record that fact without assuming it caused later delay.

Transfer, jurisdiction and panel composition

Was the case transferred between courts or prosecution offices because of jurisdiction or another recorded reason? Did panel composition change, and did domestic procedure require the hearing or trial to restart? Were the reasons recorded, and what identifiable delay followed?

Preparedness of parties and actors

What did the record show about the preparedness and availability of the prosecution, defence, experts, witnesses and other actors for scheduled hearings? Were delays attributed to late submissions, incomplete reports, missing witnesses, counsel availability, prosecution readiness or institutional co-ordination?

Use of procedural and technological tools

Were lawful procedural or technological tools, such as status conferences, video-link, witness scheduling, written submissions, severance or other case-management measures, considered and used where necessary, proportionate and compatible with fair-trial safeguards? The analysis should not assume that greater use is inherently better.

Appeal, retrial and downstream effects

Did appellate proceedings identify recurring and avoidable reasoning or procedural defects that required remittal, retrial or other rework and materially extended the case? Appeals and remedies should not be coded as inefficiency merely because they add time.



Inquiry area



Guiding questions

Positive efficiency practice

Did the court use lawful and proportionate measures that advanced the case while protecting rights? Examples may include effective planning, clear hearing schedules, disciplined handling of motions, timely rulings, witness sequencing, appropriate use of video-link, or proportionate responses to conduct that a court found abusive.

Efficiency analysis should also avoid three common errors.

1. **First**, do not equate duration with inefficiency. Serious and complex proceedings often require time. Ask whether the time taken was justified by the circumstances of the case or reflected identifiable and avoidable friction.
2. **Second**, do not equate activity with progress. A case may have many hearings and still make little progress if those hearings repeatedly fail to take evidence, resolve motions or prepare the next stage.
3. **Third**, do not treat efficiency as superior to fairness. Efficiency cannot justify curtailing preparation, restricting participation without lawful grounds, weakening required reasoning, or bypassing safeguards. Efficient proceedings proceed responsibly, not merely quickly.

An efficiency finding should identify the source, procedural stage, relevant indicator or event, the reason recorded for delay or productivity, claim level, denominator where applicable, and remaining uncertainty.

For **example**, at hearing level: In the hearing of [date], the case was adjourned because the expert report was not ready. No substantive evidence was taken. This is a hearing-level observation with potential efficiency significance linked to expert-evidence preparation and should be followed across the case trajectory.

At case level: During the proceedings, four hearings were adjourned because expert evidence was not ready or could not be discussed. The adjournments affected trial continuity and contributed to the overall case duration. The case material supports a case-level efficiency concern linked to expert-evidence planning, but does not, by itself, show whether the issue recurs across comparable cases.

At cross-case level: Across [n] of [N] eligible monitored multi-defendant corruption trials, adjournments linked to expert evidence recurred at trial and contributed to longer proceedings. The pattern was most evident in cases involving financial evidence. The evidence supports a cross-case efficiency concern and raises questions for Explain about expert availability, court control, party preparation and institutional co-ordination.

At systemic level: Across the monitored pool of serious corruption cases, repeated delays linked to expert financial evidence appear at trial and are associated with long intervals between hearings, unproductive hearings and weak sequencing of financial evidence.

Additional independent or meaningfully complementary evidence - for example, court scheduling records and expert-assignment protocols across the defined courts/institutions - shows that the same expert-assignment and scheduling workflow governs the preparation and introduction of financial expert evidence throughout the institution. Taken together, the monitored recurrence and the independent evidence of that common workflow support a systemic efficiency finding within that stated scope, specifically at the stage where expert evidence is prepared and introduced. Because the monitored pool was purposively selected, this is not a prevalence estimate for all serious corruption cases. The finding does not establish whether the driver lies in expert availability, remuneration, court scheduling, prosecution preparation, defence conduct, legal design or institutional co-ordination.

Positive efficiency findings should be stated with the same discipline. For example:

In [case], the panel used an early status conference to identify contested evidence, schedule witnesses in sequence and set dates for expert testimony. Subsequent hearings advanced the case as planned, with limited adjournments and reasoned responses to party motions. This supports a case-level good practice finding on proactive judicial case management, subject to comparison with similar cases before any broader conclusion is drawn.

2.3 Capacity

Capacity concerns the demonstrated ability of justice actors and institutions to perform the functions required by the cases they handle. In serious organized crime and corruption cases, this includes the lawful and coherent construction, presentation, management and reasoning of complex proceedings. Capacity is not simply individual knowledge; it may also depend on organizational systems, co-ordination, resources and legal enabling conditions. Analyse records, observable performance, and directly documented conditions. Explain tests why any gap persists.

The practitioner does not assess whether an individual judge or prosecutor is generally competent. The capacity lens is not a personnel or disciplinary assessment and does not replace appellate review. It asks whether the monitored record shows that an institution, office, court, panel or professional role performed the functions required for the case type. The inquiry can be organized around four components:

01 <i>Component</i> Professional / technical capability <i>Question</i> Observable quality of legal reasoning and drafting, and lawful use of specialized instruments such as financial investigation, special measures and confiscation.	02 <i>Component</i> Organizational capacity and resources <i>Question</i> Staffing, workload, expert support and case-management systems, where independently documented.	03 <i>Component</i> Process and co-ordination capacity <i>Question</i> The quality and timeliness of inter-institutional hand-offs from investigation to prosecution, prosecution to court and first instance to appeal.	04 <i>Component</i> Legal / institutional enabling conditions <i>Question</i> Whether the applicable rules and institutional design enable or constrain the required practice.
---	---	--	--

The four components provide the analytical frame; the sections that follow apply that frame to the functions performed by the principal justice actors in serious and complex cases. They are not separate checklists, and each component will not be equally visible in every case. The primary entry point is observable professional performance in the monitored record; organizational, co-ordination or enabling conditions should be incorporated where the evidence directly supports them, rather than inferred from the performance gap itself.

a) Capacity of the prosecution

Prosecution capacity concerns the ability to construct, present and sustain a serious case coherently, lawfully, objectively and on an evidence-based foundation. This may depend on professional capability, but also on the organizational support, inter-institutional processes and legal conditions within which prosecutors work. The analysis begins with what is observable in the case record and incorporates those wider conditions only where they are evidenced. It includes individualized charging, disclosure and fair-trial obligations, and the proportionate use of coercive or negotiated tools. The focus is not on whether the prosecution ‘wins’, but on whether it performs its function in a way that permits the case to be tested fairly and effectively.

In serious and complex cases, this may include the clarity and legal accuracy of the charges; the structure and presentation of evidence; compliance with disclosure and legal requirements; witness and expert preparation; lawful use of investigative, financial and negotiation tools; and the connection between the factual theory and the legal elements of each charged offence.

A core capacity question is therefore:

Did the prosecution present a coherent and legally intelligible case theory, comply with its legal obligations, and connect the charges to the evidence on which it relied?

This inquiry may be structured as follows.

Inquiry area

Guiding questions

Case theory and prosecutorial strategy

Was the factual and legal theory of the case presented coherently? Could the court, defence and observer understand the conduct alleged, the legal elements engaged and the evidence on which the prosecution relied? The practitioner should assess clarity and structure, not whether the theory was persuasive on the merits.

Charging accuracy and fact-element mapping

Did the indictment comply with domestic requirements and inform the accused in sufficient detail of the nature and cause of the accusation, including the factual basis alleged and its legal characterization, as required by the applicable fair-trial standard? Were factual allegations linked to the elements of each charge and individual defendant sufficiently to permit the case to be understood and tested? Were amendments required because the original framing was unclear or incomplete?



Inquiry area



Guiding questions

Evidence strategy

Did the prosecution explain how the evidence related to the alleged facts and legal elements, to the extent required by domestic law or the approved review criteria? Did judicial or appellate reasoning identify material gaps in the structure, legality or presentation of the evidence? The practitioner should not decide independently determine whether the evidence was sufficient to establish guilt or whether the evidentiary mix was substantively persuasive; record judicial or appellate findings and assess structure, presentation and reasoning only against the defined legal-review criteria.

Expert and financial evidence

Was expert evidence presented clearly and connected to the prosecution's stated theory? Where financial investigation, asset tracing, seizure or confiscation was legally relevant, what did the record show about whether those tools were considered, requested and integrated? Do not infer that non-use was deficient without the applicable legal and factual basis.

Witness evidence and protection

What did the record show about Steps taken to secure witness attendance, present witness evidence and seek protective measures where legally available and necessary? Were the reasons for not calling a material witness visible? Were protective measures requested and implemented consistently with defence rights?

Legality and reliability of evidence

Was prosecution evidence excluded or found unlawful, and what reasons did the court give? Does the issue appear isolated or recur in relation to collection, authorization, disclosure or presentation? Analyse the recorded legal findings; do not independently redetermine admissibility or reliability.

Use of immunity and co-operation arrangements

Did any immunity, co-operation or similar agreement comply with domestic law and the stated criteria? Were its material terms, obligations and benefits visible and reasoned where required? Where co-operation formed part of a plea agreement, did the record address the co-operation expected and the consequences of non-compliance?

Examination and cross-examination

Were examination and cross-examination structured, relevant and connected to the stated case theory? Did the prosecution respond coherently to defence challenges and evidentiary issues visible in the record? Avoid judging advocacy style where no defined standard applies.

Plea agreements

Did a plea agreement comply with domestic law, contain a factual basis where required by domestic law, and address sentence, fine, confiscation, co-operation or other measures where relevant? Did the court provide the reasoning required by law and protect the parties' rights?

The practitioner should state prosecution-capacity findings carefully. For example:

The indictment in [case] set out the alleged conduct and legal qualification but did not clearly link specific factual allegations to each element of the charged offence. This supports a case-level capacity concern regarding charging accuracy and fact-element mapping.

The source of verification should always be identified. The issue may arise from the indictment, a hearing observation, a party submission, a court remark, an evidentiary ruling, a judgement, an appellate decision, or the End-of-Case Assessment. Each source differs in appropriateness, quality and independence, and therefore in the kinds of claims it can responsibly support.

b) Capacity of the court

Court capacity concerns the demonstrated ability to manage proceedings, apply the law, maintain order, protect fair-trial conditions, give reasons where required, and bring serious and complex cases to a lawful and intelligible conclusion. As with prosecution capacity, this performance may depend not only on professional capability, but also on the organizational support, procedural arrangements and legal conditions within which the court operates.

The question is not whether the practitioner agrees with the decision, but whether observable conduct and reasoning meet the applicable standards for managing and deciding the type of case before the court. A core capacity question is therefore:

Did the court manage the proceedings, apply the relevant law and give the reasons required to allow the case to be tested fairly, efficiently and coherently?

This inquiry may be structured as follows.

Inquiry area

Guiding questions

Application and interpretation of law

Did the court apply the relevant legal standards clearly and with the reasons required by domestic law? Where domestic procedure provides for review or confirmation of an indictment, did the responsible judge, panel or other body address the applicable requirements and give reasons where required? Analyse what the record shows; do not substitute the practitioner's view of whether the indictment should have been confirmed.

Reasoning quality

Were decisions clear, coherent and non-contradictory? Could the reader follow the chain of reasoning? Did the court address material submissions and the relevant domestic, constitutional, European or international standards where applicable? Assess the quality and reviewability of the reasoning, not whether the practitioner would have reached the same result.



Inquiry area



Guiding questions

Courtroom management

Did the court maintain order and manage interruptions, party conduct, public reactions or witness intimidation in a lawful and even-handed manner? Did it intervene where conduct threatened a witness's safety, dignity or ability to give evidence, while preserving the parties' rights?

Case management and leadership

Did the court manage the proceeding actively and lawfully by sequencing evidence, addressing bottlenecks, setting realistic schedules and using planning measures where appropriate? Did the record show avoidable fragmentation?

Management of parties' motions and objections

Did the court handle motions, objections and procedural requests consistently with the applicable law and in an even-handed manner? Did it distinguish the legitimate exercise of procedural rights from conduct that a reasoned decision identified as abusive or disruptive?

Witness protection and well-being

Did the court take appropriate and lawful measures to protect witnesses where necessary? Were measures explained to participants where required, and were protection, defence rights, effective participation and transparency properly balanced?

Use of legal and procedural tools

Where tools such as video-link, protective measures, scheduling orders, severance, confiscation powers or evidentiary-management measures were legally available and relevant, did the court use or decline them lawfully, for stated reasons where required, and with necessary and proportionate safeguards? Greater use is not inherently better.

Resilience and conduct under pressure

In a high-profile case that has attracted significant media attention, did the court maintain procedural discipline, neutrality and reasoned decision-making? Record observable conduct and public statements; do not infer that media or institutional pressure affected the court without distinct evidence.

Appeal durability

Did appellate review identify recurring reasoning defects, procedural errors or misapplication of law? Did any remittal or reversal identify a first-instance issue that warrants cross-case analysis? An appeal outcome is not, by itself, proof of weak capacity.

Some capacity questions overlap with fairness or efficiency, as the same practice may engage more than one dimension. Identify the dimension most directly engaged and note any secondary implications without duplicating the finding.

For example, refusing defence time to respond to new evidence without legally sufficient grounds primarily engages fairness, while the quality of the reasoning may also raise a capacity question. Repeated failure to sequence witnesses or expert evidence primarily engages efficiency, while the case-management practice may also raise a capacity question. A judgement that fails to connect facts, evidence and legal elements may primarily engage capacity and have fairness or appellate implications.

A court-capacity finding may read:

In [case], the court issued several material procedural rulings orally. The applicable law required reasons, but none were discernible from the hearing record or any subsequent written act. This supports a case-level capacity concern regarding reasoned procedural decision-making.

Or, at cross-case level:

Across [n] of [N] eligible monitored serious corruption cases, the judgements contained limited reasoning on confiscation, even though financial gain formed part of the prosecution's case theory. The evidence supports a cross-case capacity concern regarding confiscation reasoning and the integration of financial evidence. It raises questions for Explain about legal design, evidentiary presentation, judicial practice and institutional support.

2.4 The strategic use of legal and procedural instruments: a cross-lens strand

Serious and complex cases frequently involve financial investigation, confiscation, special investigative measures, mutual legal assistance, witness protection, expert evidence, plea agreements, immunity arrangements and other instruments. Their use is a cross-cutting issue: knowledge and integration may strain capacity; timing and co-ordination may affect efficiency; and legality, proportionality and safeguards may affect fairness.

The mere existence of a tool does not establish effective practice, and greater use is not inherently better. Analyse asks whether the record shows that actors considered, used, challenged or declined the tool lawfully and for reasons appropriate to the case, and whether the necessary safeguards and reasoning were in place.

For **example**:

- A financial investigation may be legally available and relevant, but the record may show that it was not pursued;
- Special investigative material may require close review of the applicable legality and fair-trial safeguards and of the court's recorded findings on reliability or evidentiary sufficiency; whether corroboration or other safeguards are required depends on the legal context and the court's reasoning, including any special safeguards governing covert, anonymous or entrapment-related material;
- Where the record shows a concrete protection need, non-use of legally available witness-protection measures may warrant inquiry; do not infer a deficiency without identifying the applicable duty, the actor competent to seek or order the measure, and the relevant facts;
- A plea agreement may resolve the case but leave material questions about its factual basis, co-operation, confiscation, proportionality or required reasoning;
- Confiscation may be mentioned but not supported by the reasoning required under domestic law.

2.5 Reading the dimensions together

The same observed practice may engage more than one dimension. A weak indictment may raise capacity concerns and, where the record links it to amendments, disputes, retrial or prolonged exposure to unclear charges, may also have efficiency or fairness implications. Poor hearing management may raise capacity and efficiency concerns and, in some circumstances, affect equality of arms. Weak reasoning may engage capacity and fairness and may reduce the reviewability of the decision. Repeated adjournments may raise an efficiency concern and prompt Explain to test questions about preparation, witness management, scheduling or procedural responsibility.

 Observed practice	Capacity reading	Possible efficiency implication	Possible fairness implication	 Hand-off to Explain
Weak fact-legal element linkage in indictment	The prosecution theory of the case may be unclear or incomplete.	Amendments, disputes, retrials or acquittals may add avoidable time where the record links them to the defect.	Defendants may face prolonged exposure to charges that are not framed with the clarity required by law.	What directly observed conditions accompany the weakness, and which possible drivers should Explain test?
Repeated unproductive hearings	Case-management practice may be inconsistent or weak.	Trial continuity suffers.	Equality of arms or reasonable-time concerns may arise.	At which decision point does the bottleneck appear: summons, scheduling, party conduct or court control?
Thin confiscation reasoning	Reasoning on illicit gain and the handling of financial evidence may be limited.	Where inadequate reasoning creates uncertainty about the basis or scope of the order, it may complicate review or enforcement, or contribute to avoidable appeal, remittal or further proceedings. Illicit gain may remain unrecovered after lengthy proceedings.	The decision may be difficult to understand or review, which can raise fairness or reviewability concerns depending on the applicable standard.	What conditions may account for the gap between available confiscation powers, the evidentiary record and the final decision?

2.6 Good practices are findings too

Analyse should identify good practices as well as concerns. Good practices matter because they show that the system may already contain workable and potentially replicable approaches. They are harder to articulate than simple compliance: meeting a legal minimum is not automatically good practice, and a preferred outcome does not prove that the process was sound. This further matters for Recommend because reform proposals are stronger when they build on functioning practice already visible within the system.

3 Bound the finding

3.1 Bounded inference

Once the practitioner has set the claim level and applied the relevant interpretive lens, the next task is to present the analysis as a finding. Monitoring exercises are vulnerable at this point: the material may be rich and the concern real, but an unstructured conclusion may still appear selective, impressionistic, or stronger than the evidence allows.

Bounded inference is the disciplined move from observation, through significance against a stated standard, to a conclusion. The finding is the output of that reasoning. Analyse must therefore separate what happened, why it matters, and what it may indicate. The first is observation; the second is normative or operational significance; the third is inference, stated at the appropriate confidence.

For **example**:



Observation

What does the record show, directly?

In four hearings, proceedings were adjourned because expert financial evidence was not ready.



Significance

Why does the gap matter?

The adjournments affected trial continuity and raised an efficiency concern in the management of complex financial evidence.



Inference

What it may indicate?

The pattern may indicate a weakness in expert-evidence planning, court control over evidentiary timelines, party preparedness or institutional co-ordination.



Bounded finding

What does the record show, directly?

In four hearings, proceedings were adjourned because expert financial evidence was not ready.

This formulation is stronger than simply stating: *The court failed to manage the case efficiently.*

The shorter statement sounds direct, but it obscures the reasoning and assigns responsibility before the evidence has been tested. The structured version identifies the observation, explains its significance, leaves possible explanations open, and states the finding at the appropriate level.

This discipline is especially important when analysing final case outcomes. Do not code conviction, acquittal, partial conviction, sentence, confiscation, appeal outcome or retrial as success or failure.

Ask what the record shows about the pathway to the result: what was charged, what the court found established, which evidence it assessed, how it reasoned, which safeguards were observed and which legal tools were used.

3.2 Confidence language

Confidence language is not a stylistic choice; it records the strength and limits of the evidence. A useful test is whether another careful reader could reach the same conclusion from the same evidence and stated standard.

If so, the inference is likely to be bounded. If the conclusion depends on unstated contextual knowledge, the analysis needs more evidence, clearer reasoning or a narrower claim.

Use **confidence language** deliberately and apply the terms consistently across the Guide:

- The evidence **shows**: The claim is directly supported at the stated level by sources fit for the claim, independently corroborated where required, and not displaced by material contrary evidence.
- The evidence **indicates**: The claim is supported at the stated level, but one material limitation remains or independent corroboration is incomplete.
- The evidence **suggests**: The pattern is partial or emerging and is consistent with the proposed conclusion but is not yet sufficiently corroborated.
- The evidence **raises a question about**: The material is important but insufficient for a finding and should be carried forward as a hypothesis or evidence gap.
- The evidence **does not allow a conclusion on**: The available record cannot responsibly support the claim, and the boundary must be stated.

Confidence language does not create a new taxonomy that must be assigned to every finding; rather, the terms above enable the practitioner to express evidentiary strength.

3.3 Comparability

Comparability governs the choice of denominator, the construction of comparison groups, disaggregation, and every claim that a pattern persists. Compare like with like by defining the eligible group and considering offence type, defendant status, procedural route, court level, case complexity, legal tool, procedural phase, outcome, and relevant period. Do not combine contested high-status corruption trials with low-level plea cases without a reasoned analytical basis.

Where variables, such as defendant status, charge gravity, or media salience, also helped construct the monitored pool, disclose the selection and do not treat associations within the pool as estimates of the underlying caseload. Where the unit is defendant-charge, account for observations nested within defendants and cases.

3.4 Analytical safeguards

Apply these safeguards consistently to all analytical tasks.

1. Use **neutral case outcome language**: Record 'conviction', 'acquittal', 'partial conviction', 'sentence', 'confiscation ordered/not ordered' or 'appeal upheld/reversed/remitted' without coding the result as success or failure.

2. **Separate the case outcome, reasoning and pathway:** The case outcome records what the court or procedure produced; the reasoning records how the decision was justified; and the pathway examines how the indictment, evidence, hearings, legal qualification, procedural management and appeals interacted over time.
3. **Test rival explanations:** A pattern may be consistent with indictment quality, evidentiary gaps, lawful exclusion of evidence, party strategy, judicial reasoning, appellate standards, legal design, expert availability, investigative limits, or procedural safeguards. The analysis should keep credible alternatives open unless the evidence supports narrowing them.
4. **Triangulate before making a broader claim:** Compare the primary record with hearing notes, indictments, judicial and appellate decisions, procedural chronology, statistics, and, where methodologically appropriate, consultations. No source should be used to support a claim beyond what its appropriateness, quality, independence and limitations permit.
5. **Preserve uncertainty:** 'The record does not allow a conclusion', 'indeterminate at the current unit' and 'requires comparison across cases' are valid analytical outcomes. A recurring unresolved issue may justify a Deep Dive only if it also meets the materiality, feasibility and proportionality criteria.

4 Structure the finding

Bounded inference limits what may be claimed. To make the reasoning inspectable, structure each material finding around four elements: the issue or practice under examination; the applicable standard; the evidence and analysis; and the bounded conclusion. This is a presentation discipline, not a separate analytical method.

→The **issue or practice** is the precisely defined pattern, concern or good practice under examination. For example, repeated postponements in complex cases, limited reasoning for confiscation, inconsistent judicial control of plea agreements or recurring difficulty in securing witness attendance.

→The **applicable standard** provides the reference point against which the issue is assessed. Label its authority: binding domestic law; a treaty or fair-trial standard; an institutional mandate; a policy or reform commitment; or non-binding good practice. These categories do not carry equal authority, and the language of compliance must match the authority invoked. A finding may draw on more than one standard, but the basis should remain explicit.

→The **evidence and analysis** test the material against that reference point. Ask what the hearing notes, legal acts, case reports, procedural chronologies, statistics or contextual sources show; whether the issue is isolated or recurring; whether it appears across comparable cases; and whether it engages fairness, efficiency, capacity, or more than one dimension.

→The **bounded conclusion** is the finding that follows. One case may support a case-level finding; recurrence across comparable cases may support a cross-case pattern; and a pattern that meets the systemic threshold may support a systemic finding. Additional source types do not automatically strengthen the conclusion unless they are independent and relevant to the claim.

For example, the analysis should not jump from 'five hearings were postponed because witnesses did not appear' to 'the court lacks capacity'. The four-part structure makes the missing reasoning visible. The issue is repeated postponement. The applicable standard may be reasonable-time and lawful case-management requirements. The evidence and analysis ask what the record shows about service of summons, available case-management measures, party readiness, witness protection and alternatives. The bounded conclusion may be that the monitored cases indicate a recurring vulnerability in securing witness attendance and managing consequential postponements. Explain must test which contributing conditions, if any, are supported by the evidence.

Together, this structure and bounded inference produce an analytical statement that can be traced back to the evidence and carried forward to Explain. The finding should be robust enough to support explanatory testing, yet disciplined enough not to prejudge the explanation. Set the observation window

5 Route the finding

By this point, Analyse has done three things. It has set the level of the claim, read the evidence through the relevant lens, and structured the finding so that its source, standard, inference, and limits are clear. The remaining question is practical: where should the finding go next?

Routing is not a new analytical method. It is the exit map for work already completed. It prevents two opposing errors: leaving a meaningful pattern trapped in individual case records or prematurely pushing a weak observation into the reform cycle.

The route depends on the evidence currently available. Findings may remain at the hearing or case level, become cross-case patterns with a defined institutional scope or systemic findings, require further collection, trigger a Deep Dive, or move to Explain for diagnosis. Good practices follow the same evidentiary discipline and must be rights-compatible before they are proposed for protection, scaling or standardization.

The routing question is therefore:

Given the claim level and scope, evidence base, recurrence, corroboration, contrary evidence and remaining uncertainty, what is the appropriate next use of this finding?

The following table provides the practical exit map.

 What the evidence supports	 Route	 What happens next
Single hearing- or act-level observation, or evidence that supports no broader conclusion	Hearing-/act-level observation or case-level finding, depending on the unit supported	Record the observation or finding at the unit the evidence supports, with its source and boundary. It may inform internal analysis, illustrate a pathway or be followed across the case trajectory, but it carries no wider claim.
Repeated issue within one proceeding	Case-level trajectory finding	Treat this as a case-level finding followed through the Case Report and, where available, the End-of-Case Assessment. It may later support cross-case analysis if similar pathways appear elsewhere.



What the evidence supports



Route



What happens next

Issue recurs across comparable cases, but appears linked to one actor, chamber, court unit or localized practice

Cross-case pattern with localized scope

Report the finding within its demonstrated scope and route it to Explain if diagnosis is needed. Do not widen the finding without further evidence.

Issue recurs across comparable cases and different decision-makers, but corroboration for a broader claim is incomplete

Cross-case pattern; confidence or coverage limited

Record the cross-case pattern at the level supported, report n/N, missing observations and the applicable comparison group, and state the confidence or coverage limitation. Identify any additional corroboration needed before making a broader systemic claim.

Issue recurs across comparable cases and is sufficiently supported, but the current unit of analysis is too coarse to resolve the pathway

Deep-dive trigger

Scope a higher-resolution inquiry only where the issue is material, tractable, proportionate and authorized. The unit may shift, for example, from case to defendant-charge, judgement to reasoning segment, or outcome to procedural pathway.

Issue recurs across comparable cases, is sufficiently supported for the claimed scope, and evidence shows that it reflects a common rule, workflow, routine or institutional practice

Systemic finding

The finding travels to Explain with an explicit scope statement, evidentiary base and limitations. Specify whether the scope is a court, case type, procedural phase, institution, jurisdiction or region; do not describe the monitored sample itself as the system.



What the evidence supports



Route



What happens next

Good practice clears the same evidentiary gates

Good practice at the supported level; candidate for transfer where warranted

Carry the practice forward as a candidate for protection, scaling or standardization. Explain should test the conditions that enabled it, its rights compatibility and the limits of transferability before Recommend converts it into a proposal.

Evidence could not be obtained, verified or compared

Evidence gap

Record the gap in the evidence-gap register, identify the claim or comparison it affects, and route the need to targeted collection, re-scoping or a Deep Dive where proportionate. Do not automatically treat absence or inaccessibility as a finding about practice.

Routing preserves the boundary between Analyse and Explain. The handover must preserve the claim level and scope rather than force inflation to reach the next Step. Repeated adjournments linked to expert financial evidence may support a bounded cross-case efficiency finding, but they do not establish whether the driver lies in expert availability, court control, party preparation, remuneration, legal design or institutional co-ordination. Explain tests that question unless the uncertainty is sufficiently material and granular to justify a Deep Dive first.

► Before you move on



ETHICAL FRAME NOTE

Objectivity. The practitioner's professional judgement must be transparent, impartial and consistently applied. Apply the same criteria across comparable cases and record the comparison group. Record contrary evidence and unresolved disagreement. Disclose conflicts of interest, and recuse where personal, professional or financial interests could reasonably call independence into question. Never use language imputing competence, motive or improper influence without evidence directed to that proposition: the subject of a finding is a practice, not a professional. No finding leaves this Step without proportionate quality assurance covering source fit, claim level, applicable standard, contrary evidence and language, with material disagreements and their resolution recorded.

Do no harm. Before a finding is shared outside the organization, assess the risk of prejudice to ongoing proceedings, re-identification in a small pool, retaliation against justice actors, victims or witnesses, and political misuse. A bounded, well-sourced finding is not immune to contestation, but it is much harder to dismiss or misuse.



COMMUNICATION CARD

At this Step, communicate the analytical status of what you know, when someone asks what you are seeing at whatever level Analyse has established, in the language that level supports. If Analyse says the evidence suggests, a briefing cannot say it demonstrates because the meeting needs a strong headline. An emerging signal is described as one: we are seeing recurring examples and are testing whether they form a broader pattern. A useful rhythm is finding, boundary, next question — what the evidence shows, where the pattern is not uniform, and what remains open. The authority or impatience of the audience does not move a claim between categories. Once a pattern is stated clearly enough to be understood, other actors may provide material capable of testing it. They may tell you that the cases all pre-date a reform, identify a group of institutions where the pattern does not occur, point to decisions missing from the dataset or provide contrary examples. That is useful precisely because Analyse is designed to update judgements when the evidence changes.



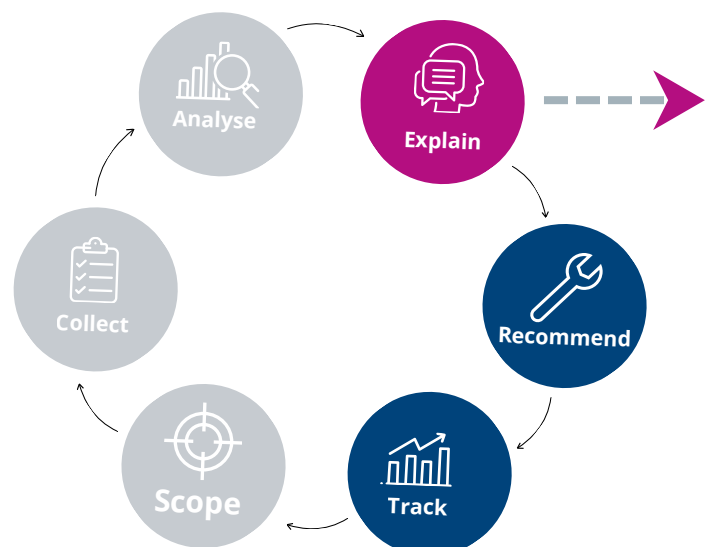
ADAPTATION NOTE

Adaptation is not merely a matter of changing examples. Practitioners should retain the analytical discipline while re-specifying the legal standards, inquiry areas, units of analysis, available sources, observable decision points and source limits for the new field. If monitoring extends beyond adjudication, the courtroom and the judicial act may no longer be the primary observation sites. When adapting Analyse, the inference discipline is unchanged: set the claim level, read the evidence through the relevant lens, bound the finding and route it. What changes is the specification.



STEP-CHECKLIST

- ☐ Is **every finding stated at the level and scope the evidence supports**, namely, hearing/legal-act, case, cross-case or systemic - and is any actor, institutional or procedural scope stated separately?
- ☐ Have you **recorded contrary or qualifying evidence** and any unresolved disagreement?
- ☐ Have you **recorded what cannot be concluded and why**, including any evidence gap that should reopen Scope or Collect, any material issue that may justify a Deep Dive, and any publication or do-no-harm risk?



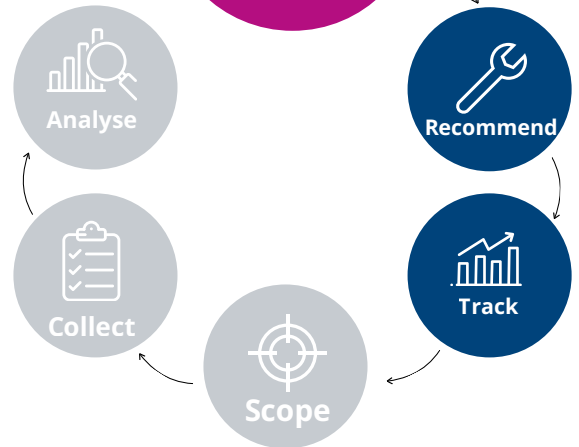
Step 4



Explain

► Where you are in the cycle

Analyse converts the evidence into bounded findings and patterns. Where necessary, a Deep Dive may strengthen earlier work by exploring a narrower question at higher resolution. Explain begins only when there is a validated, bounded finding or pattern with a traceable evidence base and clearly defined limits. It then develops and tests plausible contributing conditions: where the pattern arises, how those conditions may operate, why the practice may persist, and how strongly the evidence supports the account.



► Why this Step is important

Explain prepares the basis for reform design by identifying the conditions that plausibly contribute to the pattern, the decision points and practices through which they operate, and one or more potential entry points. Recommend then assesses those entry points for legality, rights implications, whether an institutional mandate is possible, feasibility, likely contribution, and unintended effects.

► What happens if we are not intentional about it

Without a disciplined explanatory Step, recommendations become generic. A delay finding becomes “improve efficiency”; a reasoning finding becomes “train judges”; a confiscation finding becomes “strengthen asset recovery”. Each may be relevant, but none identifies the conditions that must change or explains why the current practice persists. An intervention may deliberately be a temporary measure to reduce an immediate problem rather than a durable reform; if so, it should be labelled and assessed as such. Without that distinction, the system may appear resistant to reform when the intervention has treated the wrong condition or has addressed only the symptom.

► What you produce

By the end of this Step, the practitioner should have:

the bounded finding, applicable standard, level of claim, scope of generalization, source provenance and material contrary evidence, with a clear distinction among the observed pattern, proposed driver, working hypothesis and evidence gap, an actor-and-authority map, one or more potential entry points, the unresolved evidence gaps, and a clear routing decision to Recommend or to return to an earlier Step.



THE STEP AT A GLANCE

- 1 Turn the finding into an explanatory question**
What established pattern now needs to be explained?
 Preserve the finding's provenance, level of claim, contrary evidence and limits. Convert it into a question about the conditions that may produce or sustain the pattern.
- 2 Read across the seven diagnostic families**
Where might the pattern be produced?
 Legal design; institutional structure, mandates and resources; co-ordination; professional capacity, tools and operational support; incentives, feedback and performance pressure; professional norms; and the external environment
- 3 Explain pattern persistence**
Why might the pattern remain even after formal reform or institutional attention?
 Examine perceptions, ways of working and professional norms, and institutional structures.
- 4 Articulate competing explanatory accounts**
What else would fit these facts?
State how each proposed driver would contribute to the pattern and what else should be visible if the account is correct.
- 5 Test and narrow the explanatory accounts**
What evidence supports, qualifies or contradicts each explanation?
 Use targeted consultations, document review, existing case-trajectory analysis, Deep Dives, comparative context and media analysis to add nuance.
- 6 Assign explanatory status and route the explanatory account**
How strongly does the evidence support the proposed driver or contributing condition and who could act on it?
 Record supporting and contrary evidence, remaining alternatives, the actor-and-authority map, potential entry points and unresolved gaps. Then route the account to Recommend or back into the engine.

► What Explain does: from finding to an explanatory account

A finding is an evidence-based statement about practice. Explain receives the bounded finding together with the source material used to support, qualify or contextualize it, including hearing observations, legal acts, administrative data, consultations, media material and any Deep-dive outputs. Explain may synthesize those materials, but it must preserve the identity, independence, date, access limits and evidentiary function of each source, because each source supports different propositions. Sources that repeat the same underlying account do not constitute independent corroboration.

What changes here is the logical operation required of the evidence. Analyse reviews the evidence to establish which pattern is supported and how it relates to the applicable standard. Explain extends that reasoning by testing which conditions may contribute to the pattern and how they may operate. This is a more exposed inference because contributing conditions are rarely directly observable and must be assessed through multiple sources, credible alternatives, and explicit limits.

At this stage, the engine develops a bounded, structured explanatory account of a supported pattern. The account must preserve the scope established in Analyse – including the cases, institutions, actors and decision points to which the pattern can responsibly be generalized – and must keep the pattern distinct from isolated error. “Explanatory account” is used in a bounded sense: the Guide performs diagnostic and contribution analyses, identifying conditions that plausibly contribute to a pattern and are sufficiently well supported to inform proportionate decision-making in Recommend. It does not and need not prove strict causation. Temporal sequence, stakeholder agreement and a coherent narrative do not, by themselves, establish causation.

Monitoring may support analytical generalizations, within a stated scope of generalization, about patterns that recur across the observed cases, courts, actors or decision points. This is distinct from prevalence. The monitored pool may show that a practice recurs without establishing how widespread it is across the national caseload. Explain must state the scope of generalization and keep that distinction clear.

The intended movement is therefore not a single leap from symptom to cure but a disciplined chain of Steps:

bounded analytical finding → explanatory question → diagnostic-family scan → provisional system stability analysis, where relevant → competing or combined explanatory accounts → distinguishing evidence → testing and narrowing → final explanatory status → actor-and-authority map → potential entry points → routing.

This is where the policy engine lays the groundwork for reform design. Explain identifies a supported explanatory account and potential entry points; Recommend selects and designs the response.

1 Turn the finding into an explanatory question

1.1 Begin with a bounded analytical finding

Analyse made the reasoning beneath each finding explicit and inspectable. Explain begins only when the finding or pattern has passed the preceding quality checks: the level of claim is clear; the applicable standard is identified; the evidence base is traceable; contrary evidence and quantitative denominators, where relevant, are recorded; and the conclusion does not exceed the monitored pool. If those elements are missing, the issue returns to Analyse.

The analytical conclusion may be:

In the monitored cases, the case material indicated potentially recoverable proceeds, but the records did not consistently show that the competent authorities considered, sought or courts ruled on provisional or final asset-recovery measures, despite the material indicating potentially recoverable proceeds permitted by domestic law.

1.2 Formulate the question

Explain then asks: Which conditions may be contributing to this pattern, at which decision points, through which actors or institutional processes, and under what legal or operational constraints? How may the pattern be sustained? What evidence would distinguish among the plausible accounts?

The practitioner should not ask: “Why is the system inefficient?” or “Why is capacity weak?”, because these formulations already contain an interpretive answer. The question should retain the specificity achieved in Analyse, even while remaining open-ended. Candidate explanations may include an inadequate legal framework, limited access to financial expertise, inconsistent interpretation or practice, delays in mutual legal assistance, lawful differences in procedural posture, case complexity or a data artefact. These are hypotheses to be tested, not a menu from which the practitioner selects by intuition.

Trial monitoring and related evidence can support bounded diagnostic and contribution claims, but they will rarely establish strict causation and never an individual actor’s motive. Explain therefore asks which contributing conditions are supported and which alternatives remain open.

1.3 Set the reasoning guardrails

Explanatory reasoning begins by keeping more than one plausible account open. The practitioner does not assume that the first, most familiar, or most politically salient explanation is correct, nor does the practitioner define the problem as the absence of a preferred solution. For each account, the practitioner should state what else would be expected if it were correct and what evidence would narrow or disconfirm it. The diagnostic families generate candidate accounts; they do not determine the answer.

1.4 Set a provisional explanatory status

Explain uses four explanatory statuses to indicate how far the evidence supports a proposed account: strongly supported driver, plausible driver, working hypothesis, or evidence gap.

A provisional status may guide further testing, but the final status is assigned only after competing accounts have been tested, the system stability analysis completed where relevant, and the quality-assurance or challenge process required by the organization's governance arrangements has been completed. These statuses give practical form to the Guide's assessment of evidentiary support; they are not a numerical confidence scale.

1. **Strongly supported driver.** The evidence provides convergent support sufficient to establish, at the stated explanatory level, both the specified condition and the proposed pathway by which it contributes to the occurrence, recurrence or persistence of the pattern. Independent or meaningfully complementary sources should ordinarily be used where the nature, breadth or risk of the claim requires them. Material contrary evidence has been identified and addressed, and any residual uncertainty is stated. This label does not imply that the condition is necessary, sufficient or the sole cause.
2. **Plausible driver.** The evidence supports a specified condition and pathway, but support is incomplete, source independence is limited, the account applies only to part of the pattern, or a credible rival remains open.
3. **Working hypothesis.** A specified account that fits some of the evidence but remains insufficiently tested. The record sets out what evidence would confirm, narrow or disconfirm it. It may justify further targeted collection. Recommend may also consider whether a clearly labelled hypothesis supports a low-risk, reversible learning measure; it does not justify public attribution or irreversible reform.
4. **Evidence gap.** The pattern is visible, but no explanatory account can yet be responsibly advanced. The record identifies whether the gap concerns the absence, quality, accessibility, comparability or resolution of evidence. The output names the precise missing information and what should be done to obtain it, including routing the issue to targeted completion, a Deep Dive or a recorded unresolved gap.

The explanatory statuses must be applied with discipline, including in terms of:

- **Terminology:** As causality requires caution and proper evidence, “driver” and “contributing condition” carry the ordinary work of Explain; “cause”, “produced”, “determined” and “effect” are reserved for exceptional cases where the evidence genuinely supports them.
- **Proportionality:** Explain records the explanatory status and its evidentiary limits. The status records how far the evidence reaches, and nothing more. It does not describe how convinced the practitioner privately feels, nor what should be done about the pattern. Recommend then decides whether that status is sufficient for a particular response, taking account of the stakes, cost and reversibility of the decision the account will inform. A less strongly supported account may support a low-risk learning measure; structural change, public attribution or irreversible action requires stronger evidence and review.



Practitioner Note

An explanatory account need not eliminate every possible explanation before it can inform reform, but it must be strong enough to inform a proportionate option appraisal in Recommend. It must disclose credible alternatives and support the assigned explanatory status. Explain does not decide that an intervention is justified; it supplies Recommend with the strongest account the evidence can responsibly support and the limits that must shape any response. Avoid reading the effectiveness dimensions specifically as drivers. A Capacity finding does not imply a “professional capacity” explanation; it may be explained by legal design, incentives, co-ordination, resources, norms or external pressure. Likewise, Fairness or Efficiency findings may have capacity-related drivers.

2 Read across the seven diagnostic families

2.1 Diagnostic families context

The bounded findings from Analyse identify patterns in practice. In the EBTM application, those findings will often rest substantially on trial-monitoring evidence: hearing observations, legal acts, case trajectories and cross-case patterns. That evidence can show where and how a problem becomes visible and may provide initial clues about contributing conditions. It will not ordinarily establish internal workflows, motives, resource-allocation decisions or pre-indictment processes without additional sources.

For example, a monitored hearing may show that financial evidence was repeatedly unavailable, but not why the relevant expertise was not secured. A judgement may fail to explain the link between facts, evidence and legal elements, but not reveal whether this arose from unclear guidance, workload pressure, limited supervision, fragmented responsibility or formal or perceived pressure to file or dispose of cases quickly. A recurring pattern of adjournments may identify the immediate recorded reason for delay without revealing the administrative routine, resource constraint or co-ordination failure that contributes to recurrence.

The seven diagnostic families (a synthesis of lessons from RTM implementation, legal-institutional analysis, organizational analysis and public-administration diagnostics) are proposed as a practical search grid for generating and organizing plausible explanations, not a validated, exhaustive or causal taxonomy. Not all seven need to be investigated exhaustively in every inquiry, and they do not authorize the practitioner to answer questions that the evidence base cannot support but rather provide a structured device for generating and organizing possible explanations.

The families are intended for multidisciplinary use. Trial monitoring contributes evidence of observable practice, while legal experts, policy advisers, data analysts, programme managers and other relevant profiles may contribute the specialist knowledge needed to test different parts of the explanatory account. No single practitioner is expected to conduct every part of the inquiry alone.

In this sense, they can also be viewed as prompts that prevent professional reflex from becoming an explanation. A lawyer may see a legal gap. A trainer may see a skills gap. A manager may see a workflow problem. A donor may see weak ownership. Any of these may be correct, but professional instinct alone does not meet diagnostic standards. Placing a finding within a family identifies a line of inquiry; it does not establish that the family contains the driver. Their purpose is simply to widen the explanatory inquiry, identifying evidence needs and structuring multidisciplinary testing. In a snapshot:

 Diagnostic family	 Core question	 Evidence to examine
Legal / normative design	How, if at all, does the rule or legal route materially contribute to uncertainty, contradiction or impracticability?	Legal framework; jurisprudence; reasoned decisions; recurring interpretation at a defined decision point.

**Diagnostic family****Core question****Evidence to examine**

Institutional structure, mandates and resources	How are authority, ownership, governance, staffing, funding or system-level support aligned with the function?	Mandates; budgets; staffing; governance and escalation routes; actor-and-authority map.
Co-ordination and information flow	Is relevant information, responsibility or action lost at a hand-off within the observable justice pathway?	Process maps; timelines; transmission records; case files; interfaces between actors or systems.
Professional capacity, tools and operational support	Can actors and teams perform the task with the knowledge, guidance, tools, expertise and support available?	Outputs and case practice; access to experts, precedent, templates, systems, supervision and peer review.
Incentives, feedback and performance pressure	What behaviour is formally or informally rewarded, tolerated, discouraged or left without feedback?	Evaluation criteria; targets; promotion and discipline systems; management messages; data use; perceived and actual incentives.
Working routines and professional norms	Do established ways of working reproduce the pattern even where formal rules permit another practice?	Repeated workflows; templates; local variation; peer expectations; longitudinal and multi-source evidence.
External environment and narrative pressure	In what way, if at all, do relationships and pressures outside the immediate case-processing chain shape risk, discretion or institutional behaviour?	Executive and legislative relations; governance bodies; security threats; donor conditions; bar and civil-society dynamics; media and public narratives.

a) Legal / normative design

Some patterns may reflect the legal or normative framework. The practitioner should distinguish an absent, contradictory or impracticable rule from lawful discretion, evolving jurisprudence, different factual or procedural postures and uneven implementation. The problem may lie in criminal procedure, substantive offence definitions, asset-recovery rules, publication rules, plea-agreement safeguards, appellate standards, limitation periods or institutional mandates.

A claim that legal design contributes to the pattern should be grounded in authoritative text, implementation history, relevant appellate interpretation and evidence linking the rule to the observed decision point. Reports that “the law is unclear” are a lead, not a conclusion. The analysis should test whether the rule materially contributes to the pattern or whether implementation, guidance, professional practice or another condition provides the stronger account.

Example: If competent authorities apply asset-recovery measures inconsistently and the domestic framework leaves unclear who must initiate the measure, the applicable evidentiary threshold, the protection of bona fide third parties, or the route from provisional restraint to final confiscation and enforcement, legal design may contribute. If the framework and appellate guidance are clear, the practitioner should test implementation, capacity, co-ordination and incentive explanations rather than assume a legislative gap.

Useful **guiding questions** include:

1. What binding domestic law, procedural rules and applicable standards govern the decision point, and are they identifiable and internally coherent?
2. Are divergent interpretations tied to the wording or structure of the rule, or do they reflect different facts, procedural postures or lawful discretion?
3. Is the procedural route complete from initiation through to decision and enforcement?
4. Does the framework impose a duty without identifying the competent actor, authority, information, safeguards or workable procedure needed to perform it?
5. Does compliant or better-functioning practice under the same framework indicate that the problem can be addressed without legal amendment?
6. Is the variation within legitimate judicial discretion or evolving jurisprudence, and what authority supports that conclusion?

b) Institutional structure, mandates and resources

Some patterns may reflect a mismatch between the function the system expects and the authority, ownership, governance, staffing, funding or system-level support available to perform it. The inquiry should identify the structural condition and avoid attributing poor performance to an individual actor without separate evidence.

This family includes mandate fragmentation, unclear ownership, weak escalation routes, insufficient staffing, unstable funding, lack of specialized support, missing governance mechanisms, absent quality-control routines and responsibility gaps between policy and practice. Resources belong here when they concern the system’s structural capacity to perform the function: for example, specialized prosecutors, financial experts, court staff, interpreters, IT systems, asset-management structures, data officers, publication staff and realistic time allocations.

This differs from professional capacity. A prosecutor may understand how to develop an asset-recovery application under domestic law, but the function may still fail if financial-analysis support is unavailable, lawful access to asset information is delayed, case allocation leaves insufficient time for complex preparation, or the prosecution office lacks an appropriate quality-review mechanism. In that situation, institutional structure and resources may contribute alongside individual or team capability.

Example: If final decisions repeatedly identify defects in the linkage between alleged facts, evidence and offence elements, but no lawful routine feeds those reasons into pre-filing review, case-strategy supervision or professional learning, the explanatory account should examine institutional feedback and quality-review arrangements rather than assume an individual skills deficit.

Useful **guiding questions** include:

1. Which actors hold formal authority at each relevant decision point?
2. Which actor initiates, approves, reviews, transmits or enforces the action?
3. Where does responsibility become shared, fragmented or unowned?
4. Are staffing, time, funding and specialized support proportionate to the mandate?
5. Is there an escalation route when one part of the chain does not act?
6. Do governance arrangements and safeguards preserve judicial and prosecutorial independence and autonomy while still supporting lawful accountability and organizational learning?

c) Co-ordination and information flow

Serious and complex cases may depend on information and action moving accurately and on time from investigation to prosecution, prosecution to court, first instance to appeal, judgement to enforcement, provisional restraint to asset management and final confiscation, and operational data to management review.

A pattern may arise at a hand-off even where each actor performs their part of the task. Relevant material may arrive late or in an unusable form; appellate guidance may not reach lawful quality-review or training processes; publication systems may not transmit decisions consistently; or performance data may use incompatible identifiers or definitions. The practitioner should identify the precise hand-off, the information or action expected, the applicable legal or procedural duty and the point at which the transfer fails.

Example: If hearings repeatedly fail to proceed because notifications, detainee transport, expert availability or scheduling information are not co-ordinated, the pattern may reflect a hand-off or scheduling-system problem rather than an inference about judicial willingness.

Useful **guiding questions** include:

1. What information or action must pass from one actor or stage to another?
2. When and in what format is it transferred?
3. Does the receiving actor have the authority and capacity to use it?
4. Are delays or losses visible at the same interface across comparable cases and can they be distinguished from a substantive deficiency arising earlier in the process?
5. Do case identifiers, data systems or terminology prevent the information from travelling?

d) Professional capacity, tools and operational support

This family concerns practical ability to perform. Depending on the actor and the material lawfully observable, it may include knowledge of complex offences, financial investigation, digital evidence, asset recovery, admissibility and reasoning requirements, drafting, case management, expert evidence and

review. It should not infer guilt or a skills deficit from the case outcome alone.

Useful **guiding questions** include:

1. What professional task does the pattern require actors to perform?
2. Is the expected standard clear and accessible?
3. Are guidance, templates, precedent and expert support available and usable?
4. Have actors received relevant training, and is there evidence that the training was applied?
5. Could workload, legal uncertainty, information gaps or incentives explain the same output better than a knowledge or skill deficit?

e) Incentives and performance pressure

Some patterns persist because the performance environment rewards activity rather than effect. Institutions may reward indictment volume without comparable attention to legal sufficiency, count completed hearings without examining avoidable adjournment, record training attendance without assessing use, or report convictions without examining appeal durability, lawful sentencing or asset recovery. Any assessment of judicial or prosecutorial performance must respect decisional independence and prosecutorial autonomy and avoid outcome quotas or pressure towards conviction.

Formal adoption or reported activity may therefore coexist with unchanged practice: a rule is adopted, a strategy approved, training delivered or a system installed, but the evidence does not yet show that the relevant routine changed.

Example: If prosecutors are evaluated by number of indictments rather than quality, weak cases may continue to enter the system because the incentive favours volume over case construction.

This family also examines feedback and organizational learning: whether the institution collects reliable data, reviews relevant outcomes and uses the results to adjust guidance, support or workflow. The absence of such a loop may contribute to recurrence, but the practitioner should specify the information that was available, who was expected to use it and what evidence shows that learning did or did not occur.

Useful **guiding questions** include:

1. What do formal evaluation, promotion, discipline, budget and reporting systems measure?
2. What messages do managers and professional-governance bodies send about priority and acceptable risk?
3. Are incentives and disincentives actual, perceived or both?
4. Do speed, volume or closure targets conflict with quality, fairness or durable outcomes?
5. Are actors rewarded for using specialized tools or penalized for the additional time they require?
6. Does the institution collect, analyse and lawfully use information from appeals, dismissed or amended charges, acquittals, delays, complaints and reform implementation to adjust practice without treating a particular lawful outcome as a performance failure?

f) Working routines and professional norms

Formal rules do not operate in isolation. Repeated workflows, templates, peer expectations and informal practices shape how work is performed. Explain should examine these practices without treating an institution as having a single culture or motive.

This family may include highly formalized or minimum-compliance routines, risk-averse decision-making, template reasoning, reluctance to depart from established practice, or professional expectations about what counts as complete work. More sensitive labels, such as self-censorship, resistance or avoidance of controversial decisions, require specific evidence and should not be inferred from a small number of unpopular outcomes.

This family also examines how institutions absorb reform. A recommendation may be accepted, a template completed, a publication rule formally applied, or a case-management tool installed without evidence that the underlying workflow changed. The practitioner should distinguish recent or partial implementation from stable non-use or deliberate avoidance.

Example: If courts hold scheduled hearings but, where domestic procedure permits, do not use case-management powers to narrow issues, manage evidence, set lawful timelines or address recurring causes of adjournment, the account may point to an established working routine rather than an absence of legal authority.

Useful **guiding questions** include:

1. What counts as acceptable or complete work among relevant professional peers?
2. How do practices vary across courts, offices, teams or locations?
3. What happens when an actor departs from the established way of working?
4. Is the new practice recent, piloted, unevenly implemented or formally adopted without operational absorption?

g) External environment and narrative pressure

Serious OCC cases often unfold in a contested information environment. Public narratives may frame a case as accountability, impunity, selective justice, political revenge, institutional weakness or elite protection before the court has reached a final decision.

Media and contextual analysis do not prove what happened in court or why a legal decision was made. They may provide evidence about public narratives, documented threats, reputational conditions and perceived independence. Any claim that the external environment affected institutional conduct requires separate, source-specific support.

Example: If a case that has attracted significant media attention is tried under intense public pressure to secure a conviction, fairness risks may arise. If a comparable case is almost invisible in the media despite high public interest, the absence of scrutiny may also be analytically relevant. In both situations, media is not proof of the legal issue. It is evidence of the environment in which the legal issue is processed.

Useful **guiding questions** include:

1. What external actors have lawful authority over resources, appointments or institutional priorities, and what safeguards protect judicial independence and prosecutorial autonomy?
2. What public or professional narratives surround the issue?
3. Are threats, retaliation risks or public statements documented and independently corroborated, and is there evidence linking them to institutional conduct beyond temporal association?

2.2 The evidence required may sit outside the courtroom record

Some diagnostic questions can be examined through the evidence already assembled in Collect and Analyse. Others concern matters that are not directly observable through trial monitoring, including:

- institutional mandates and ownership;
- internal workflows and approval routes;
- staffing, budgets and workload allocation;
- access to specialist expertise or operational tools;
- performance indicators and promotion incentives;
- informal professional expectations;
- perceptions of risk and responsibility;
- co-ordination between institutions.

Where these matters are relevant, Explain should first determine whether the existing evidence base already contains a reliable source. It should seek additional evidence (as guided in the table below) only where that source can distinguish between accounts or materially affect the explanatory status, the actor-and-authority map, the stability analysis or the routing decision.

 Explanatory uncertainty	 Appropriate source or method
Whether the pattern recurs across a wider or more specific set of cases, actors or decisions	Existing cross-case analysis; if insufficient, a Deep Dive
How a pattern operated through the trajectory of a particular completed case	End-of-Case Assessment from Analyse, with outcome-bias safeguards
What formal mandates, workflows, standards or performance arrangements require	Targeted document review, verified against implementation evidence
How actors understand the routine, its risks and the practical constraints surrounding it	Structured consultations; direct evidence of reported perspectives, not case facts
Whether administrative distribution, resources or performance patterns support an explanation	Official and administrative data with definitions, denominators and coverage limits



Explanatory uncertainty



Appropriate source or method

Whether public narratives or reputational pressure shape the operating environment

Media and contextual analysis; not proof of influence without additional evidence

Whether another institutional arrangement shows a different way of handling a comparable difficulty

Comparative contextual analysis after testing legal and institutional comparability



Method Note

Consultations are especially important when the inquiry concerns reported routines, risk perceptions, informal norms or institutional hand-offs that cannot be observed directly. They may provide primary evidence of professional perspectives or reported experience, but do not by themselves prove case facts, prevalence or a causal mechanism.

2.3 Reading across diagnostic families

Most patterns may involve more than one contributing condition. The purpose of the diagnostic families is not to force the pattern into a single box, but to prevent the practitioner from stopping at the first plausible explanation. The account should compare the relevant families, identify which combination is best supported and state the scope and limits of that conclusion. Recommend, not Explain, assesses actionability.

For **example**:

The available records indicate that asset-recovery measures were not consistently considered, sought, ordered or enforced in monitored serious corruption and organized crime cases in which the case material indicated potentially recoverable proceeds.

Read through the seven families, the same pattern may reveal different possible drivers:



Diagnostic family



Possible drivers

Legal / normative design

Domestic law may be unclear about which authority initiates provisional or final measures, the applicable evidentiary standard, the treatment of third-party interests, the availability of extended or non-conviction-based routes, or the relationship between confiscation, sentencing, civil recovery and enforcement. Any explanation must be grounded in the applicable domestic framework and authoritative interpretation.



Diagnostic family



Possible drivers

Institutional structure, mandate design and resources	Responsibilities for tracing and preserving assets, financial investigation, applications or charging, adjudication, enforcement and asset management may be distributed across several bodies. The relevant question is whether domestic law assigns each function clearly and whether funded, staffed and rights-compliant arrangements connect the stages.
Co-ordination and information flow	Asset or financial information may not move through the case at the required time or in a usable form. Investigators or financial specialists may identify assets late; prosecutors may lack usable analysis; courts may not receive the applications, evidence or information required under domestic law; or enforcement and asset-management bodies may receive incomplete or non-executable orders. The account should identify the precise hand-off rather than describe a general failure.
Professional capacity, tools and operational support	Relevant actors may lack the specialized guidance, financial-analysis support, templates or review needed to connect alleged conduct, potentially recoverable proceeds, ownership, evidence and the applicable legal basis. The record should identify which actor had the relevant task and what the domestic law required at that stage.
Incentives and performance pressure	Formal or perceived performance arrangements may give greater visibility to indictments, convictions or case closure than to lawful recovery and enforcement. This account requires evidence of the relevant measures, messages or reporting practices and must avoid pressuring prosecutors or judges towards outcomes.
Working routines and practices	Asset recovery may be treated as an exceptional add-on rather than integrated, where domestic law permits, into early case planning, adjudication and enforcement. The practitioner should distinguish a stable routine from lawful case-specific reasons not to pursue a measure.
External environment and narrative pressure	Public and media attention may focus on conviction or acquittal rather than on whether proceeds were lawfully restrained, confiscated, recovered or returned. This may make asset recovery a performance issue, but it does not explain why a competent authority acted or failed to act.

The explanatory task is not to list all seven and call that an explanatory account, but to determine which condition or combination is best supported and which alternatives remain open.

For example, if the evidence base shows that competent authorities did not consistently consider or seek asset-recovery measures, financial analysis was often unavailable or late, judicial acts did not address recoverable proceeds where domestic law required or permitted consideration, and institutions did not track enforcement, the monitored asset-recovery gap appears plausibly associated with a fragmented institutional pathway across investigation, prosecution, adjudication, enforcement and asset management, reinforced by uneven financial-analysis support and reporting arrangements that give limited visibility to recovery outcomes.

The account remains bounded to the monitored settings and does not establish that any actor deliberately avoided confiscation. It identifies the relevant decision points, shows why training alone may be insufficient and frames the question Recommend must appraise: which lawful condition or combination of conditions should be addressed, in what sequence and with what safeguards, to improve asset-recovery practice?



Practitioner Note

In politically sensitive work, the practitioner should keep credible rival explanations open. Legal, evidentiary, organizational, resource, co-ordination and complexity factors should be tested alongside evidence of improper pressure, differential treatment or intervention. No explanation is presumed correct. The task is to compare competing accounts against the traces each should leave in the evidence and revise the assessment as that evidence develops.

Capacity and structural conditions matter in two directions. Weak procedures, fragmented authority, poor information flows or limited professional support may contribute directly to poor performance, while also creating openings for pressure and making improper intervention harder to detect or resist. Conversely, strong institutional capacity can act as a safeguard against influence: clear rules, professional competence, transparent decision-making, effective oversight and resilient institutional routines can reduce opportunities for interference and strengthen actors' ability to withstand it. **Capacity and influence are therefore neither mutually exclusive nor simply competing explanations.**

3

Explain why the pattern persists

Identifying a plausible driver is not always enough. Where the pattern has continued despite efforts to address it, including formal reform, institutional attention or repeated recommendations, Explain should also examine what keeps reproducing the practice despite those efforts. This is a secondary inquiry applied to the potential accounts, not a separate diagnostic framework.

The practitioner should prepare a system stability account. Its function is to test the rules, incentives, resources, risks, dependencies and working practices that may make the current practice easier, safer or more workable than the alternative. At this point the analysis is provisional and must be revisited after competing accounts have been tested.

Identifying a plausible driver is not always enough. Where the pattern has continued despite efforts to address it, including formal reform, institutional attention or repeated recommendations, Explain should also examine what keeps reproducing the practice despite those efforts. This is a secondary inquiry applied to the potential accounts, not a separate diagnostic framework.

The practitioner should prepare a system stability account. Its function is to test the rules, incentives, resources, risks, dependencies and working practices that may make the current practice easier, safer or more workable than the alternative. At this point the analysis is provisional and must be revisited after competing accounts have been tested.

Not every finding requires such an analysis. Use it where a pattern has persisted, recurred or survived a relevant attempt to change it; not where the pattern is recent, local, temporary or has not yet been meaningfully addressed.

A persistent practice may reflect several different conditions: incomplete or uneven implementation; formal compliance without practice change; a lawful or rights-protecting function; perceived professional or institutional risk; resource or authority constraints; local variation; or deliberate obstruction. These possibilities are not equivalent and must be demonstrated rather than assumed.

The inquiry should also distinguish individual perspectives from formal mandates, routines and institutional conditions. An institution should not be treated as a single actor with one motive; nor should institutional behaviour be reduced to isolated individual choices. However, repeated interactions among actors, rules, incentives and routines can generate stable system dynamics that persist even where no individual intends the overall result.

Persistence is therefore not evidence of “resistance” by default. Ordinary implementation difficulty, principled caution, lawful safeguards and deliberate obstruction are analytically different states that the practitioner must distinguish among them.

After the seven diagnostic families’ analysis has produced potential accounts, the practitioner applies the following three prompts only to the account or accounts that remain plausible, to parse out a system stability account. Together these prompts ask what makes the current practice workable, risky to depart from or difficult to replace; whose perspective supports that account; what contrary evidence exists; and what would have to change for an alternative practice to become viable while preserving legality, rights and institutional independence.

3.1 Perceptions and decision conditions

Actors may make decisions under constraints of time, information, authority and role. This prompt examines how those conditions shape the perceived risks and feasibility of acting differently. Neither the framework nor the evidence permits the practitioner to infer that a particular actor was anchored, loss-averse or resistant. Perceptions are not directly visible in a judgement or case outcome and must be explored through appropriately designed consultations, implementation records, surveys or other relevant evidence.

Useful **guiding questions** may include:

- How do relevant actors describe the problem and the expected change?
- What risks do they associate with acting differently, and what evidence supports that account?

- Who bears the immediate cost of change and who receives the benefit?
- Are actors uncertain about what is expected or about how their performance will be judged?
- Do different professional groups perceive the same reform differently?

3.2 Ways of working

At the level of working practices, the practitioner examines how shared routines, informal expectations and professional norms shape what counts as acceptable practice. A few cases or interviews may suggest a line of inquiry; they do not establish how an institution “thinks”.

Useful **guiding questions** may include:

- Which practice is repeated even where another lawful route is available?
- What templates, workflows or peer expectations sustain it?
- What is the professional or social cost of being the first to work differently?
- Does the new practice fit or challenge the professional identity of the group?
- Is adoption recent, partial, uneven or limited to formal compliance?
- What examples of functioning alternative practice already exist within the system?

3.3 Institutional structure

At the institutional level, the practitioner asks how mandates, resources, accountability arrangements, data systems, incentives and decision rights affect the relative feasibility and perceived risk of the current and alternative practices. Actors may understand and support a reform but be unable to act differently because the structure does not provide the required authority, information, resources or safeguards.

For example, reporting lines may discourage escalation; evaluation systems may emphasize volume over quality; budgets may not align with mandates; relevant data may be unavailable; managers may lack authority; or interdependent institutions may each wait for another actor to move first. Each proposition requires evidence from the relevant mandate, workflow, data or actor account.

Judicial independence and prosecutorial autonomy should not be treated as performance barriers to be engineered away. They are legal and normative constraints on both explanation and reform design. The account should distinguish legitimate decisional independence from organizational arrangements that impede lawful accountability, learning, support or co-ordination.

Useful **guiding questions** may include:

- What formal authority, information, resources and support are required for a different practice?
- Who can act, who must approve, and who can block or delay the action?
- What lawful independence or rights-protecting function does the current arrangement serve?
- What prior reforms were attempted, and what happened during implementation?
- Did reform alter the formal structure without changing the operating conditions?
- What evidence would show that the practice is beginning to change?

Together, the three prompts should produce one system stability account, attached to the surviving explanatory account or accounts. Its purpose is to identify the functions, risks, dependencies and constraints that Prescribe must take seriously if a reform is to alter the conditions reproducing the existing practice, rather than merely adding another rule, tool or process on top of them.

For example, judgements may continue to provide limited reasoning after training because of workload, template design, limited research support, weak feedback, uncertainty about appellate requirements or professional norms that prioritize formal completeness over explanatory depth. Any such assessment must remain anchored in domestic law and applicable reasoning standards, not disagreement with the outcome. If the combined account is supported, it may indicate that additional training alone is unlikely to address the pattern. Explain should identify any lawful or rights-protecting function that a later response must preserve, without selecting the replacement intervention.

4 Articulate competing explanatory accounts

The diagnostic-family scan produces a longlist of candidate drivers. The practitioner must now turn that longlist into a manageable set of rival, or combined explanatory accounts. The accounts should be specific enough to test, and each should identify the proposed condition, the pathway through which it may contribute, the cases or decision points it could cover, and what else should be visible if it is correct.

Competing accounts need not be mutually exclusive. Serious patterns often arise from interacting conditions. The discipline is to avoid an unstructured list in which every family is named, and none is weighed.

For each candidate explanation, the practitioner should state:

- The **proposition** – what condition, mechanism or institutional arrangement may be contributing to the pattern.
- The **expected trace** – what else should be visible if that explanation is correct.
- The **supporting evidence** – which existing sources point towards it.
- The **contrary evidence** – which observations, cases or sources do not fit it.
- The **distinguishing evidence** – what additional information would help separate it from rival explanations.
- The **scope** – which cases, actors, stages or institutional settings the explanation could reasonably cover.

This turns explanatory reasoning into a test. If unclear legal design is contributing to the pattern, the practitioner should expect to find ambiguity or contradiction in the applicable framework, divergent interpretations arising from that ambiguity, or repeated uncertainty at the relevant decision point. If limited professional support is contributing, the record may show a lack of expertise, unavailable templates, insufficient review, weak access to precedent, or recurrent difficulty with the same technical task. If improper pressure is contributing, the practitioner would need evidence of pressure, unexplained deviation, differential treatment, or another trace that cannot be inferred merely from an unpopular outcome.

An explanation should be narrowed or discarded when its expected traces are absent, when contrary evidence is stronger, or when another account explains the pattern more coherently. Several explanations may remain. Serious institutional patterns may reflect interacting conditions rather than a single isolated cause.

5 Test and narrow the explanatory accounts

The diagnostic families and the provisional system stability analysis generated candidate drivers, and the previous move organized them into competing explanatory accounts. Those accounts do not validate themselves. The next task is to identify the evidence that would distinguish among them and to test the propositions proportionately.

No single method is reserved exclusively for one type of uncertainty. A question about recurrence may be addressed through ordinary cross-case analysis or may require a Deep Dive. Consultations and document reviews are often combined. An End-of-Case Assessment may inform more than one question. Comparative context may challenge an explanation without establishing the domestic cause.

5.1 Match the source to the claim

Each source supports different propositions. Hearing observation may show observable courtroom conduct but not private deliberation. Judicial acts may show formal decisions and recorded reasoning but not unrecorded motives. Procedural histories may show sequence and recorded delay. Statistics may show distribution only where definitions, denominators and coverage are clear. Consultations may directly evidence reported perceptions, professional experience and perceived incentives. Formal documents may outline mandates and procedures, but not whether they are operational. Media may show public statements and the narrative environment, not institutional influence without additional evidence.

Collect records and preserve the sources. Explain matches those sources to specified claims, tests rival accounts and identifies what remains unsupported. The purpose is not simply to collect more information, and no source should be converted into stronger evidence by relabelling it.

However, Explain should not become an open-ended research phase. Additional evidence is justified where it can materially confirm, narrow or disconfirm a candidate explanation or affect its final explanatory status or routing.

The following sources and methods are commonly used to test explanatory accounts.

a) Structured consultations

Structured consultations serve a dual function. As described in Collect, they provide insight into professional experience, perceived incentives, informal routines and institutional conditions that observation alone may not capture. In Explain, they test specified explanatory accounts, and fill evidence gaps on professional experience, perceived incentives, informal routines and institutional hand-offs. Evidence-gathering and hypothesis-testing consultations should remain distinct in purpose, timing and record from earlier Collect sessions and later institutional-engagement or co-design sessions in Recommend.

Consultations may serve as primary evidence for claims about reported perceptions, professional experience, and perceived incentives, provided the claim is labelled accordingly, and the participant selection, methods, and limitations are transparent. They should not ordinarily stand alone as evidence for case facts, prevalence, legality, or causal mechanism.

A consultation protocol should record at minimum:

- explanatory question and hypotheses being tested;
- categories of actors whose knowledge is relevant and why;
- question guide, including prompts designed to test rival explanations;
- whether the account is first-hand, reported or inferential;
- areas of convergence, disagreement and possible institutional interest; and
- what documentary or observational evidence is needed to corroborate the account.

Consultations should invite disconfirmation. Rather than asking whether expert shortages cause delay, the practitioner may ask how expert evidence is requested, assigned, prepared and scheduled; at which stages delay arises; what other explanations are plausible; and what records would show this.

Consultations may support institutional understanding of a pattern, but they must not become an institutional veto or private re-litigation. The merits of ongoing cases should not be discussed. Completed public cases may be used, where necessary, to examine systemic routines, appellate feedback or implementation, provided the discussion is structured, based on the public record, appropriately anonymised and does not seek personal assessments of guilt or blame.

b) Targeted document review

A targeted document review is a limited inquiry into specific material needed to test an explanation within the existing scope. It may examine internal guidance, workflow rules, evaluation criteria, training materials, staffing arrangements, jurisprudence, selected decisions, process records or data definitions. Any non-public material must be obtained and handled in accordance with the applicable access, confidentiality and data-protection protocols.

The review should be tied to a precise proposition. Its purpose is not to collect everything available, but to determine whether the formal system, operational record or documented practice supports the candidate explanation.

For example, a hypothesis that quality feedback is absent may be tested through review of appellate-feedback mechanisms, internal review instructions, meeting records, training plans and evidence of how acquittal reasons are used. The existence of a policy on paper should be distinguished from evidence that it operates.

c) Deep Dives

A Deep Dive is different from a targeted document review. It is the policy engine operating at higher resolution around a narrower question. It formally re-enters Scope, Collect and Analyse where the existing pool or available source set cannot answer a material explanatory question.

The Deep Dive may confirm the pattern, narrow its scope, qualify a candidate explanation, reveal a different mechanism or return an evidence gap. It should not be used to prove a preferred theory at all costs.

d) End-of-Case Assessment

The End-of-Case Assessment is situated in Analyse, where it reads the proceeding retrospectively from the defined closure point and identifies case-specific factors that may have contributed to the pathway or outcome. Explain does not repeat that assessment. It uses the completed case-trajectory account as one source for testing whether a candidate driver is consistent with what happened inside finalized cases. One case cannot establish a systemic driver but recurrence across cases and the pathway within one case complement one another.



Method Note

The assessment is especially vulnerable to hindsight and outcome bias. The practitioner should ask: would the conduct, reasoning or omission have been assessed in the same way if the outcome were unknown or different? The record may also lack pre-indictment information, defence strategy, confidential material or the full legal uncertainty faced by the actors.

e) Comparative contextual analysis

Comparative analysis can test whether an explanation is plausible, reveal contextual variables and identify alternative institutional arrangements. A foreign or neighbouring system is not treated as a ready-made solution, and a similar pattern elsewhere does not establish that the domestic pattern has the same cause.

Comparison is most useful where it sharpens the inquiry: whether another legal route avoids the identified ambiguity, whether a different authority map closes a hand-off gap, or whether a comparable reform failed under similar implementation conditions. Its main role in choosing and adapting a response belongs to Recommend.

f) Media as contextual information

Media enters the engine at several points. Collect records relevant material and source characteristics. Analyse reads it for fairness, perception, pressure and triangulation. Explain asks what the information environment may help to clarify about institutional risk, public expectations and the conditions under which a pattern operates.

The practitioner should consider the reach, affiliation, source practice and role of the outlet or speaker. Media content is not a proxy for public opinion. Legitimate public scrutiny and criticism should be distinguished from prejudicial statements, threats or improper pressure. Temporal association between coverage and institutional conduct does not establish influence.

The relevant question is not only “What did the media say?” but “what proposition can this source support about the environment, and what further evidence would be required to connect that environment to institutional conduct?”

5.2 Record the test

The explanatory process should be auditable. For each material candidate explanation, the practitioner should complete one record before testing and update it after each material inquiry. The record should show the claim, expected and disconfirming traces, source independence and limitations, alternatives considered, evidence that supported, contradicted or narrowed the proposition, and the reason the provisional explanatory status changed or remained the same.

The task now is to assign the explanatory account its final explanatory status: strongly supported driver, plausible driver, working hypothesis or evidence gap.

The determination should reflect the account as a whole:

- How consistently the proposed driver fits the observed pattern;
- Whether it explains the relevant case trajectories and institutional practice, without exceeding the stated scope of generalization;
- Whether competing explanations have been tested and narrowed;
- Whether the persistence account is supported by evidence rather than assumed; and
- Whether source independence, material contrary evidence and unresolved gaps have been addressed and recorded.

The practitioner should assign the strongest explanatory status the evidence can support. Where several drivers reinforce one another, the explanatory status may apply to the combined account rather than to each component separately. The status states how far the evidence permits the account to travel; it does not measure the importance or urgency of the problem. Finalize the determination only after the applicable quality-assurance or challenge process, then incorporate it into the handover to Recommend.

6 Assign explanatory status and route the explanatory account

Before routing the account, the practitioner should confirm that the proposed pathway is specified; source quality, independence and limitations are documented; at least one credible rival has been considered; contrary and null evidence are visible; the scope of generalization and decision points are bounded; the applicable legal standard is correct; and any sensitive attribution has passed the required legal, methodological and security review. High-stakes, public or institution-specific accounts warrant enhanced scrutiny; the form and degree of reviewer independence should follow the organization's governance and risk arrangements.

6.1 Assemble the handover record

What moves forward is not the finding alone. Recommend needs a bounded explanatory account: the conditions that may contribute to the pattern; the decision points and practices through which they may operate; why the practice may remain stable; how strongly the account is supported; where authority and responsibility sit; which potential entry points merit appraisal; and what remains unknown.

The routed explanatory account should therefore identify:



Element



What must be stated

Finding being explained	→ The bounded analytical finding received from Analyse, including its scope and level of claim.
Proposed driver or combination of drivers	→ The condition or interacting conditions that most plausibly contribute to the pattern.
Operating mechanism or practice	→ The routine, decision point, hand-off, incentive or institutional process through which the proposed driver may contribute to or sustain the pattern.
Explanatory status	→ Whether the diagnosis is a strongly supported driver, plausible driver, working hypothesis or evidence gap.
Supporting and contrary evidence	→ The principal evidence that strengthens the diagnosis and the evidence that qualifies, weakens or challenges it.
Reason the practice remains stable	→ The evidenced functions, risks, dependencies, implementation conditions, rules, incentives, norms or constraints that may keep the practice stable.
Institutional responsibility or ecosystem gap	→ The actors with legal authority, operational or resource control, information dependencies, approval or review roles, and any shared or missing responsibility.
Unresolved evidence gaps	→ What remains unknown, what could not be tested, and how that uncertainty limits the diagnosis or the reform response it can justify.
Potential entry points	→ One or more points in the institutional pathway where a change could plausibly affect the supported mechanism, clearly distinguished from a selected intervention or a determination of feasibility.
Routing decision	→ Whether the account moves to Recommend, returns for targeted collection or analysis, triggers a Deep Dive, remains a working hypothesis, or requires a separate sensitive-matter protocol.

For example:

Finding: In the monitored serious corruption cases, case material indicated potentially recoverable proceeds, but the records did not consistently show that the competent authorities considered, sought or ruled on provisional or final asset-recovery measures at the stages required or permitted by domestic law.

Explanatory account: The pattern appears plausibly associated with a fragmented asset-recovery pathway, uneven access to financial-analysis support and performance arrangements that place greater emphasis on case disposal or conviction than on recovery of proceeds.

Operating pathway: The available evidence indicates that financial information is not consistently carried from investigation and prosecution through adjudication, enforcement and asset management. Relevant authority and dependencies are distributed across several bodies, and the monitored pathway does not show a consistently functioning end-to-end co-ordination mechanism.

System Stability: The current arrangement may reduce immediate workload or co-ordination risk for some actors. At the same time, responsibility and the costs of omission are dispersed and receive limited visibility in existing reporting systems. This account remains a hypothesis unless actor and documentary evidence support it.

Explanatory status: Plausible combined driver. The account is supported by monitored case trajectories, judicial-act review, administrative data and consultations, but limited access to internal allocation and quality-control practices leaves credible alternatives open.

Actor-and-authority map: Investigative bodies, financial intelligence or specialist services, where relevant, prosecution services, courts, enforcement bodies, and asset-management institutions hold distinct legal roles in the function. The map must reflect the domestic allocation of authority and protect adjudicative independence.

Evidence limits: The review could not fully examine internal financial investigation, case allocation, or quality control decisions.

6.2 The actor-and-authority map

The stakeholder map in the Scope baseline should now be updated to an actor-and-authority map. This is not a search for a single responsible actor. It shows how formal authority and operational responsibility are distributed across the relevant pathway. In doing so, it should identify legal mandates, operational and resource control, information holders, implementation partners, approval or veto points, accountability bodies, and affected groups. These constraints bound reform design; they are not performance obstacles to be removed.

Actor or body	Formal mandate	Operational role	Information or resource dependency	Approval / veto / review point
[Actor 1]	Legal authority	What the actor does in practice?	What it must receive or rely on? Any dependencies?	Where it can authorize, block, review or escalate?
[Actor 2]	[]	[]	[]	[]

The map should identify shared ownership, hand-offs, dependencies and ecosystem gaps without implying that an institution has power it does not possess. Where no actor owns the whole function, that absence may itself be part of the explanatory account.

6.3 Potential entry points

A potential entry point is a place in the institutional pathway where a change could plausibly interrupt, enable or redirect the mechanism identified in Explain. It may be a legal decision point, a review process, a data flow, a support function, an incentive or feedback channel, or a professional routine.

An entry point is not yet an intervention or a feasibility determination. Explain may identify that a pattern enters through pre-filing review, expert assignment, appellate feedback, scheduling information or performance reporting. Recommend must still decide whether and how to act after examining the implications for legality, fair trial, and other rights; judicial independence; prosecutorial autonomy; institutional mandate; likely contribution; urgency; tractability; cost; and political feasibility and unintended effects.

6.4 Routing decision

The route follows from the state of the explanatory account (see table below for more guidance). The practitioner should not advance every issue to Recommend: an unstable finding returns to Analyse, an obtainable source gap returns to Collect, and a question requiring a redesigned pool or higher-resolution inquiry triggers a Deep Dive.

	State of the account		Route
The starting proposition is not yet a bounded analytical finding.	→	Return to Analyse.	
A material source or case trajectory is missing but obtainable.	→	Return to Collect for targeted completion.	
The question requires a redesigned pool or higher-resolution analysis.	→	Trigger a Deep Dive through Scope → Collect → Analyse.	
Only a working hypothesis can be supported.	→	Record the hypothesis and evidence need; do not force a recommendation. A low-risk learning option may be considered later under safeguards. Where safeguards are met, the account may move to Recommend to appraise a low-risk measure.	

**State of the account****Route**

A plausible or strongly supported driver is established and the actor map and entry points are clear.

→ Move the bounded explanatory account to Recommend.

The issue concerns possible improper influence or sensitive misconduct.

→ Apply the separate escalation and reporting protocol in addition to the appropriate methodological route.

► Before you move on



ETHICAL FRAME NOTE

Explain is the most inferential and interaction-intensive part of the engine.

Methodological **objectivity** requires the practitioner to keep competing explanations open, match sources to the claims they can support, record contrary evidence, distinguish perception from fact and avoid defining the problem as the absence of a preferred solution. No inference about an individual's motive, bad faith or improper influence should be made merely because an outcome is controversial or inconsistent with expectation.

Non-intervention applies especially to consultations. The practitioner may test systemic hypotheses, ask about procedures and institutional routines, and discuss completed public examples where necessary and safe. The practitioner should not seek or offer views on the merits of an ongoing case, invite private relitigation of a decision, advise an actor how to handle a case, or use access to influence an outcome.

Confidentiality requires clear terms on attribution, note-taking, data retention and access to raw records. The practitioner should consider anonymization and the risk that a small professional group can be re-identified even where names are removed. Participants should be informed of any limits to confidentiality, including mandatory procedures for sensitive or unlawful disclosures. Raw consultation material should be accessible only to authorized personnel and retained in accordance with the applicable protocol.

Judicial independence, prosecutorial autonomy and fair-trial rights are not obstacles to be traded away in the search for institutional performance. They constrain and shape both the explanatory account and the later reform response.



COMMUNICATION CARD

At this Step, communication is about exposing an emerging explanation to institutional feedback and to competing accounts. You are testing it, not asking counterparts to validate it. Explain's own guidance on structured consultations (§4.5) already covers the basics. Communicate the explanatory account at the established explanatory status, communicate a plausible driver as plausible, not causal — and do not bury a well-supported explanation in ambiguity merely because it may be contested. Listen for information the primary evidence could not easily reveal: e.g., which informal routines override formal ones, where responsibility becomes blurred, and why available tools are or are not used. Ask what alternative explanation the institution would offer and, importantly, what evidence would distinguish that explanation from yours. Bear in mind that institutions often hear technically neutral findings as judgements about competence, autonomy or institutional worth. Identify what is actually being contested. A challenge to the evidence, sample or method is answered methodologically; a reaction to what the finding is felt to mean is not, and restating the methodology will not resolve it. Name the implication, separate it from the claim you are making, and re-anchor in the professional value the pattern touches; that is consistency, legal certainty, reviewability etc.



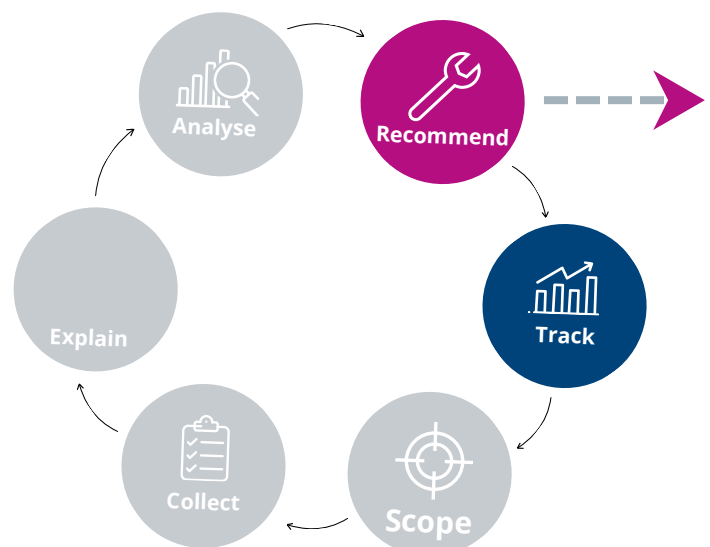
ADAPTATION NOTE

The explanatory procedure can extend beyond organized crime and corruption, but the source base, decision points, relevant driver families, and safeguards must be revalidated for each domain. The seven families are written to assess serious corruption and organized crime, so an adopter should treat them as prompts rather than transfer them unchanged. A legal-aid inquiry may require appointment, payment, and defence-preparation data; a gender-based-violence inquiry may require victim-safety, protection-measures, and service-referral evidence.



STEP-CHECKLIST

- ☐ Can you **trace the account back to** the bounded analytical finding?
- ☐ Can you see the **rival explanations** considered?
- ☐ Can you identify the evidence for and against each, and understand **why the proposed driver is given its explanatory status**?
- ☐ Can you see **why the pattern may persist**?
- ☐ Can you **confirm that the applicable legal standard** and the limits of judicial independence, prosecutorial autonomy and fair-trial rights are accurately stated?



Step 5



► Where you are in the cycle

Explain provided supported explanatory accounts and potential entry points, which Recommend receives as the basis for reform design. In this Step, the policy engine moves to proposed action that competent actors can adopt, authorize or implement. It is the most demanding Step because evidence must be translated into reform within legal mandates, rights safeguards, resources and operational realities.

► Why this Step is important

Arriving at the point of crafting recommendations follows the careful execution of the previous Steps. Following this through requires continued respect for the methodology, namely by translating a bounded explanatory account into proposed actions that are evidence-based, legally sound, institutionally legitimate, politically supportable and operationally feasible.

► What happens if we are not intentional about it

If this Step is weak, the recommendation may stray from the evidence and reflect the practitioner's or institution's prior assumptions about what should work. These assumed models may be correct, but when they are not, they lead to poor recommendations that create reform fatigue, give institutions reasons to dismiss surfaced issues, and reinforce the very stagnation the policy engine is meant to overcome.

► What you produce

By the end of this Step, the practitioner should have:

a reform recommendation package comprising one or more recommendation units, each with a main recommendation and any enabling sub-recommendations, together with a supported explanatory account; expected results; stated explanatory status, where relevant; explanations of selection decisions and logic; an identified addressee, decision authority and implementation lead, as applicable; adoption or disposition route; and handover to Track.



THE STEP AT A GLANCE

1

Construct a theory of change

What conditions must change for a different practice to become possible, likely, safe and sustainable?

Convert the diagnosed conditions into a reasoned pathway, with the assumptions, dependencies and risks it relies on stated openly.

2

Define the recommendation structure

How should the theory of change be divided into coherent recommendation units?

Set the reform objective and the recommendation target, then shape each unit around one main recommendation and the sub-recommendations that create its enabling conditions.

3

Generate and screen reform options

What lawful and evidence-based policy measures could alter the diagnosed institutional performance conditions?

Generate options against pre-set core parameters, then apply safeguard screening to reject or redesign what cannot responsibly be proposed.

4

Decide among options and draft the recommendation units

Which options are most defensible, and how do the selected units work together as one reform recommendation unit?

Choose based on evidence, lawful authority and implementation readiness, then draft each recommendation unit with a built-in tailored results' chain.

5

Assemble and test the reform recommendation package

Do the recommendation units work as one design?

Sequence them, identify dependencies, and test the package against the theory of change as a whole rather than unit by unit.

6

Choose the adoption or disposition route and hand off to Track

How can each recommendation unit realistically enter the reform environment, and what must Track receive?

Select the adoption or disposition route, name the decision authority and implementation lead, and hand over baselines, milestones and evidence sources.

► What Recommend does: a recommendation begins with the explanatory account

This is not the Step where the practitioner simply writes down what should happen next. It is the Step where the bounded explanatory account becomes a policy intervention. Recommend translates a received diagnosis into relevant and defensible potential interventions that are evidence-driven, supported by an explicit theory of change and constructed as a coherent package rather than a disconnected activity list.

Three terms organize the work that follows:

Term	 Meaning in this Guide
Theory of change	The reasoned account of how changing the diagnosed performance conditions or mechanisms is expected to alter institutional or professional practice and contribute to the broader reform objective, including the assumptions, dependencies and risks on which that pathway relies.
Reform recommendation unit	The smallest coherent block of reform design within the broader reform recommendation package. It addresses one connected reform issue or branch of the theory of change and ordinarily contains one main recommendation together with the sub-recommendations needed to create its enabling conditions.
Reform recommendation package	The selected set of recommendation units through which the overall theory of change is put into practice. It shows how the units relate, depend on and sequence with one another, and how they collectively address the conditions sustaining the various diagnosed patterns that challenge effective administration of justice as per core reform inquiry.

1 Recommendation design begins with constructing a theory of change

Explain asked why the current routine persists, while Recommend asks what would have to change for a different routine to become possible. Understanding why a routine persists means unpacking the institutional system logic: what it preserves, who bears the risk if it changes, which incentives sustain it, which formal and informal norms make it safe, and which resource or feedback gaps allow it to continue. Routines may persist for many reasons, for example, because they distribute responsibility, reduce professional or institutional risk, accommodate resource constraints, satisfy performance pressures, or preserve accepted relationships of authority.

A reform that does not account for those functions may be formally adopted while the underlying practice remains materially unchanged.

The theory of change therefore stems from the explanatory account. It identifies the desired practice, the conditions that must shift, and the sequence through which those shifts are expected to contribute to the broader reform objective. It does not yet select the concrete intervention. It establishes the logic against which possible interventions will later be judged.

To construct the theory of change, the practitioner begins with the explanatory account produced in Explain. The task is no longer to ask why the current practice persists, but to ask what would have to become different for the desired practice to become workable and sustainable.

The practitioner should reverse the diagnosis. Where Explain identified unclear expectations, missing tools, weak authority, misaligned incentives, limited capability, fragmented co-ordination, professional risk or another stabilizing condition, Prescribe asks what condition would need to replace or alter it. The focus is not yet on a preferred intervention, but on the performance conditions that must change. This matters because the same visible problem can require very different recommendations depending on the condition that is sustaining it.

To do so the practitioner must apply a diagnosis-driven recommendation design approach. This Guide recognizes that reform actors bring their own biases and conceptions of model institutions, shaped by their experiences, jurisdiction, legal tradition, institutional background, comparative exposure, and priorities. For example, a practitioner may anchor in a familiar model from another jurisdiction; a legal expert may treat harmonization as the natural cure for variation; a trainer may see every practice gap as a skills gap; or a court administrator may view delay primarily as a case-management failure.

Any of these readings may be accurate on their own merits, but that does not mean they fit the specific context or respond to the evidence or analysis. While these models may provide inspiration, they should not lead to predetermined outcomes.

To avoid simply resorting to the most familiar options when designing reforms, the practitioner should review sources of bias, including mandates, professional background, programme incentives, missing perspectives, and any actual or perceived conflicts of interest. This helps delay the selection of a remedy, allowing each identified solution to be reviewed equitably and meaningfully. Each identified solution should be treated as a hypothesis, namely, if this solution was introduced, what condition would it change, and why would that change alter the routine reflected in the finding?

Explain has already identified the supported drivers, operating pathway and any conditions sustaining the current practice. Recommend now translates those diagnosed conditions into design-relevant conditions and uses them to construct the theory of change. Do not force every diagnosed driver into the performance-condition grid below, which is a design aid for operational and professional conditions that complements the legal-authority analysis or the seven-family diagnosis in Explain.

Performance-conditions as the map to a theory of change

 Performance condition	 Recommendation question	 Possible response	 Justice-system safeguard
Information, expectations and feedback	Do actors know what is expected, how decisions should be made, and what feedback comes from later stages?	Guidance, practice notes, feedback loops, structured review, clearer standards.	Respect decisional independence; distinguish support and learning from informal appellate control.
Tools, resources and process design	Do actors have the instruments, systems, time, staffing and resources required?	Templates, workflows, staffing, expert support, registry access, process redesign.	Confirm lawful data access, defence equality, interoperability, cybersecurity and recurrent costs.
Incentives and consequences	Does the system reward the desired behaviour or reward something else?	Evaluation reform, quality criteria, managerial review, recognition, removal of perverse targets.	Do not create conviction, sanction, speed or confiscation quotas, or timeliness targets that pressure decision-makers to sacrifice fairness, reasoning or lawful discretion.
Knowledge and skills	Do actors know how to perform the task?	Training, mentoring, peer review, benchbooks, practical exercises, specialized modules.	Use only where a demonstrated capability gap exists; protect confidentiality and case-specific independence.
Assignment fit and institutional capacity	Are actors realistically able to perform the task given workload, complexity, specialization and mandate?	Specialization, workload redistribution, team-based case handling, dedicated units, realistic mandates.	Preserve transparent assignment, access to justice, defence capacity and clear escalation routes.
Confidence, perceived risk and professional norms	Do reported or observed professional norms, perceived risks or confidence affect whether actors can use the expected practice?	Leadership signalling, communities of practice, peer learning, protection from retaliation, ownership-building.	Avoid coercive signalling and unsupported inferences about individual motive; protect actors who raise lawful concerns.

The theory of change, in short, explains how an intervention is expected to alter the diagnosed condition, making the desired routine easier, safer, clearer, better supported or more rewarded than the current one.

The following three examples show the practical use of the matrix above:

Finding: Confiscation was not pursued or reasoned consistently in monitored serious corruption and organized crime cases in which alleged illicit gain was central to the conduct.

Explanatory account: A plausible explanatory account is that the pattern reflects a fragmented confiscation chain. Financial information does not travel consistently from investigation into charging, adjudication and enforcement; no actor owns the process end to end; specialist support is limited; and performance systems place greater value on conviction and case closure than on asset recovery. The practical equilibrium is stable because each institution can complete its immediate procedural task without assuming responsibility for the final recovery of assets. The additional workload and co-ordination risk are concentrated on individual actors, while the cost of failing to recover assets is dispersed across the system.

Theory of change: If responsibility for confiscation is made continuous across investigation, charging, adjudication and enforcement; if financial information is available early and in a usable form; if practitioners have the guidance and specialist support needed to connect alleged illicit gain to the legal requirements; and if feedback and performance systems treat lawful consideration, reasoning and follow-through on asset-recovery issues as part of quality serious-case handling, then, where legally and factually warranted, confiscation is more likely to be considered early, pursued consistently, reasoned adequately and carried through to enforcement. This in turn should improve the consistency with which legally recoverable illicit gain is identified and adjudicated and, where recovery is ordered, carried through to enforcement, without turning recovery rates into performance quotas.

The change pathway depends on assumptions: the legal framework provides an adequate confiscation route; relevant financial and ownership information can be accessed; institutions have authority to establish the required workflows; and enforcement and asset-management bodies can act on final orders. It also faces an absorption risk: new procedures may be adopted formally while confiscation remains peripheral to case strategy. The recommendation structure must therefore affect ownership, workflow, support and performance expectations, not merely add a template or training event.

This is what turns the problem into a solution without breaking the evidence chain.

The theory of change now provides the structure of the response, but not yet the recommendation package. It identifies the conditions that must shift and the sequence through which those shifts are expected to alter practice.

More than one performance condition may sustain the same routine. The practitioner should therefore avoid designing around the most visible condition alone. Instead, identify the material conditions that appear to work together to reproduce the current practice, distinguish those that must change jointly from those that are secondary, and identify any gateway condition without which other measures are unlikely to take effect.

Not every condition needs to be addressed through the same recommendation unit. Some may fall outside the scope, mandate or evidence base of the inquiry, or may already be addressed through complementary reforms. These should still be recorded as dependencies, assumptions or complementary reform needs, so that the package does not imply that its own measures are sufficient where they are not.

At this point, the practitioner is not yet drafting recommendations. The design target should nevertheless be clear enough to state:

This routine persists because the missing or distorted condition is [performance condition or combination of conditions]. The theory of change seeks to alter that condition through [expected change mechanism], so that [desired practice] becomes more likely and contributes to [broader reform objective].

The next task is to divide that logic into coherent recommendation units. Each unit takes responsibility for one connected branch of the theory of change; the reform recommendation package keeps the full pathway visible.

2 Define the recommendation structure

Before generating concrete reform options, the practitioner should define the changes that the recommendation package will need to produce, and how those changes relate to one another. The starting point is not the reform measure itself — training, a new procedure, or a digital tool etc. — but what that measure is expected to make possible or different in the institution.

A proposed measure may, for example, create a new capability, clarify responsibility, provide a usable tool, improve access to information, establish a co-ordination arrangement or change a working routine. These are different kinds of change and should not be treated as interchangeable simply because they can all be described as “reform implementation”.

The practitioner should therefore organize the recommendation structure around a sequence of enabling changes and the practice they are intended to support. Concrete interventions can then be generated and tested against that structure. This helps prevent the recommendation package from becoming a list of activities and keeps attention on the institutional conditions that must actually change for the desired practice to emerge and endure.

The following subsections therefore introduce the components of that structure in advance. They give the practitioner a design template to generate and assess reform options in the next section.

2.1 Define the broader reform objective and recommendation target

The practitioner should first distinguish between the broader reform objective and the recommendation target. The broader reform objective is the wider institutional or public-value improvement to which the recommendation package is intended to contribute. It sits beyond the change that any single recommendation can reasonably deliver or claim. Depending on the reform problem, this may include more effective administration of justice, stronger accountability for serious crime, greater consistency and legal certainty, improved protection of rights, or more durable justice outcomes.

The recommendation target is the more confined change that the recommendation unit seeks to bring about and that the evidence, mandate and institutional context allow the practitioner responsibly to prescribe. The target should therefore be set at the highest level of change that can be credibly supported, rather than reduced prematurely to the measures intended to produce it. Training, guidance, new procedures, co-ordination arrangements, specialist support or digital tools may later form part of the supporting architecture. They are not the recommendation target unless the evidence justifies prescribing only at that level.

The reform objective explains why the change matters. The recommendation target defines what substantive change this recommendation unit is trying to achieve. The supporting measures then address what must be put in place for that change to become possible and sustainable.

Example

The reform objective is that serious crime should not remain financially rewarding and that lawful asset recovery contributes effectively to accountability and the restoration of illicit gain.

The recommendation target may be for example, that confiscation and asset-recovery issues are identified, developed, adjudicated and followed through consistently and lawfully in relevant serious cases. This is the substantive change that the main recommendation may carry.

Early confiscation planning, structured handover of financial evidence, specialist review of confiscation requests, clearer allocation of responsibilities or mechanisms for following enforcement may then become supporting measures because they create conditions needed to realize that target.

At this stage, the practitioner is defining the target and the conditions it requires, not yet selecting the intervention through which those conditions will be created.

5.2.2 Define the recommendation unit

A recommendation unit addresses one coherent reform topic. It ordinarily contains a main recommendation and supporting sub-recommendations, arranged as an intervention logic.

The **main recommendation** carries the desired change at the level that the evidence, diagnosis and authority structure can responsibly support. It states the reform direction: what should change in institutional performance, professional practice, an enabling output or an agreed process. Institutional maturity and implementation readiness should inform the adoption or disposition route.

The **sub-recommendations** identify the concrete conditions that must be created for that change to become possible. They may address rules, procedures, tools, resources, skills, co-ordination, feedback, incentives or routines.

The unit is coherent where, if all sub-recommendations were implemented and used as intended, the main recommendation would become materially more likely to be realized in practice. The sub-recommendations should not become disconnected tasks, and the main recommendation should not carry every operational detail.

In this sense, a recommendation unit does not also need to contain every measure that might be necessary to realize in full the goal articulated in the main recommendation. It captures the branch of the theory of change that this inquiry can responsibly support within its scope, evidence base and mandate. Other actors or initiatives may work towards the same target through complementary measures grounded in different diagnostics. The requirement is therefore not totality, but coherence: the sub-recommendations must form a credible pathway through which this unit contributes to the target it names.

Each recommendation unit therefore puts one branch of the package-level theory of change into practice. A unit may address, for example, ownership of the confiscation chain, professional tools and support, or performance feedback.

Example

In the confiscation example, the recommendation target may be more consistent and lawful handling of asset-recovery issues in serious cases. The sub-recommendations in one unit may focus on early confiscation planning, internal case-handling routines and structured handover of financial evidence. Those measures may materially support the target, even though other necessary conditions — such as legislative reform, digital asset-tracing capacity, or enforcement arrangements — may sit with other actors or parallel reform tracks.



Method Note

Change happens across a spectrum, ranging from immediate procedural improvements to long-term institutional transformation. There is no rigid mathematical formula to separate them; rather, it depends on what the change potential is within the system. This is not an exact science. The discipline is to keep the formulations lean so that the implementation milestones (discussed in the next chapter) can be set to capture the appropriate level of change.

The practitioner now knows what the eventual recommendation package must contain and how its elements must relate to one another but has not yet selected the interventions that will populate it. The next task is to generate a defensible range of reform options and test which of them can best perform the functions identified in the theory of change and the provisional unit-level results chains.

3 Generate and screen reform options

The explanatory account may support more than one plausible intervention. One option may be technically stronger but politically more difficult. Another may be easier to adopt but less transformative. A third may be slower but more sustainable. A fourth may better protect judicial independence but offer weaker short-term visibility. The task is not to write the most impressive recommendation. It is to generate a defensible range of routes to change and screen out options that cannot proceed in the specific operational context.

3.1 Generate reform options

This is the point at which the practitioner must recognize that a recommendation is not merely a technical Step but a choice among possible routes to change. There are generally four core parameters to be reconciled when scoping reform options: technical soundness, political supportability, operational feasibility, and standards and safeguards.

3.1.1 Core parameters

a) Technical soundness

Does the evidence support the claim? Is the recommendation legally correct? Does it respect fair-trial standards, judicial independence, prosecutorial autonomy and institutional mandate? Does the issue require specialized expertise before the recommendation is finalized? A technically weak recommendation should not move forward merely because it is attractive or easy to communicate.

Importantly at this point certain standards and safeguards must be respected, as a first-order review, when assessing technical correctness. The following parameters test whether a proposed option is lawful, legitimately owned and operationally grounded before it is developed any further.

 Parameter	 Practitioner question
Normative basis	Which binding domestic rules, constitutional requirements, applicable international standards, authoritative interpretations or policy commitments support the recommendation?
Lawful authority	Which actor may lawfully decide, direct, adopt, fund, administer, review or support the proposed action?
Rights and proportionality	What fair-trial, equality, privacy, data-protection, victim, witness, defence and third-party interests are engaged, and is the proposed measure necessary and proportionate?
Independence and autonomy	Could the measure direct individual adjudication or prosecution, create outcome pressure, distort case assignment or blur lawful governance and decisional functions?
Confidentiality and proceedings	Could recommendation design, institutional engagement or publication disclose protected information, prejudice pending proceedings or imply guilt or bad faith without evidence?
Implementation conditions	What resources, legal amendments, procurement, data systems, expertise, sequencing or partner support are required?

Do no harm and proportionality



Are the proposed options too narrow or staged in a way that foreseeable harms are disproportionate or cannot be mitigated?

Evidence and confidence



Is the proposed option framed more strongly than the explanatory account supports?

b) Political supportability

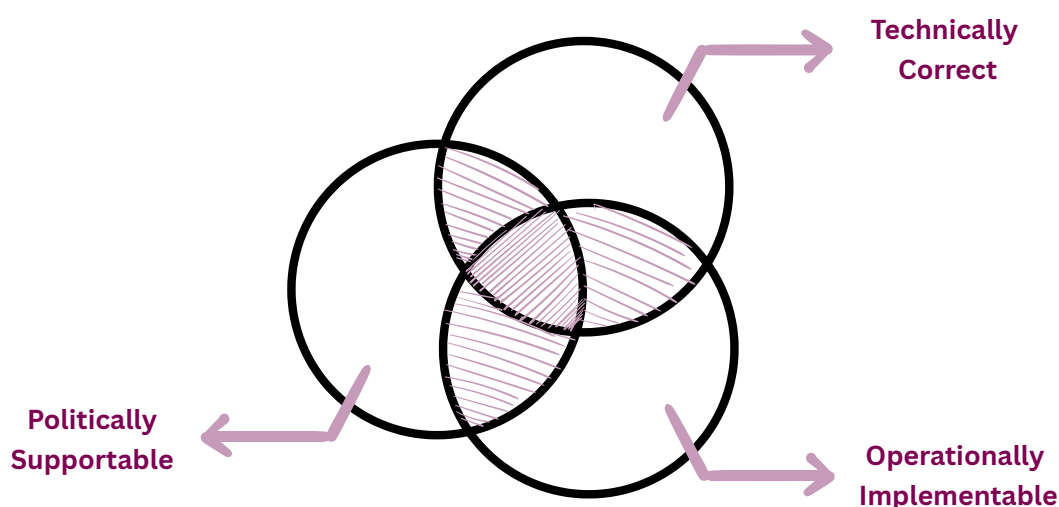
Who must tolerate, endorse or at least not block the change? Does the recommendation threaten authority, reputation, workload, resources, autonomy or political interests? Can it be framed in terms of operational quality, legal certainty, predictability, fairness, or institutional resilience rather than blame? A politically difficult recommendation may still be necessary. But its route must be designed.

c) Operational feasibility

Can the institutional recipient of the recommendation act within mandate, time, resources and administrative capacity? Is the change too broad? Is it dependent on another institution? Does it require budget, staff, IT, anonymization capacity, legal amendment, expert support or partner assistance? An operationally infeasible recommendation may need to be narrowed, staged, piloted or referred.

3.2 Deciding amongst recommendation options

These parameters should be applied together rather than sequentially. Among technically sound options (the status quo should also be included as a reference option, since retaining the current arrangement carries its own technical, political and operational implications), the practitioner compares political supportability and operational feasibility to identify the intervention with the strongest realistic pathway to implementation.



The question is then no longer only: What does the evidence show and how can we respond? It becomes: Given the diagnosis and the available options, which intervention is the most defensible, feasible and strategically viable choice? Ideally the recommended policy measure sits in the middle of the Venn diagram above but this is seldom the case in practice.

The matrix below illustrates one way of comparing reform options across the three parameters. It is not intended to produce a mechanical ranking: the relative importance of technical soundness, political supportability and operational feasibility will depend on the reform context, timing and purpose of the intervention. An option that performs less strongly across the full set may still be the better strategic choice if, for example, it can unlock a blocked process, build institutional trust, create early ownership, demonstrate feasibility or establish the conditions for a more ambitious reform later.

	TECHNICALLY CORRECT	POLITICALLY SUPPORTABLE	OPERATIONALLY IMPLEMENTABLE
POLICY 1			
POLICY 2			
STATUS QUO			

The purpose is to make the trade-offs visible. The practitioner should ask not which option accumulates the most positive marks, but which combination of strengths, weaknesses and uncertainties offers the most defensible pathway forward at that point in the reform process. Where a weakness can be mitigated through sequencing, piloting, coalition-building or complementary measures, that should form part of the design rather than automatically disqualifying the option.

The practitioner may need to consider which option best addresses the diagnosed condition, which is most likely to shift the relevant working practice, which actor has decision authority or implementation leverage, which best protects judicial independence and fair-trial safeguards, which is feasible within existing resources, which can be sequenced or piloted, and which can later be tracked. Substantive priority and implementation feasibility should be weighed against each other. A severe, high-impact issue may require engagement even when immediate implementation is difficult, and a feasible option should not be prioritized merely because it is easy.

In public reform, decision criteria are not purely technical, and different actors may value them differently. These differences should not be concealed in the recommendation. They should be recognized as part of the decision-making process. This does not remove professional judgement from the recommendation design; it disciplines it so that, even where the final choice remains partly subjective, the reasons can be traced, tested and defended.

Once the mandatory standards have been met, compare the discretionary criteria.

Three concepts guide this process.

1. First, the practitioner must **prioritize**. Not every supported explanatory account can or should become a recommendation at the same time. Priority can be assessed against three questions: how severe is the problem, how materially could change affect the reform problem, and how feasible is institutional action at this point in the reform cycle? That said, before prioritizing, the practitioner must determine what to prioritize. A serious finding that is not yet sufficiently explained should ordinarily be routed back for further Explain work rather than scored as a substantive recommendation option. A clearly labelled working hypothesis may be appraised here only for a low-risk, reversible learning or test measure designed to reduce the stated uncertainty, consistent with the routing rule in Explain. Others may be well evidenced but have limited potential to affect the reform problem. Others may be feasible but too minor to justify attention in a consolidated package.
2. Second, where mandate, confidentiality, security and non-intervention considerations permit, proposed options should be **pressure-tested** with relevant counterparts. This is not a handover of methodological control, and it is not an invitation to dilute the finding. It is a reality check on ownership, feasibility, sequencing, institutional mandate and possible unintended effects. Counterparts may confirm that the recommendation is actionable, identify a better vehicle, point to an existing reform process, or explain why a proposed route would be absorbed without changing practice. Institutional engagement can strengthen the package when it tests the intervention logic without surrendering methodological control or the evidence base.
3. Third, the practitioner must apply a **coherence test**. Each recommendation should be traced back to the validated finding, the explanatory account, the system stability analysis where relevant, and the diagnosed condition or mechanism identified as contributing to the pattern. It should then be walked forwards to the kind of evidence that would later show whether implementation has begun to change practice. The practitioner should find no broken link.

4 Decide among options and draft the recommendation units

Once an option has been selected, it can be formulated as a recommendation unit. The main recommendation should state the desired effect or change at the highest level the context can responsibly support. The sub-recommendations should identify the enabling conditions and concrete actions needed to make that change more likely.

The discipline is the same in every unit. The main recommendation should not carry every operational detail. The sub-recommendations should not become disconnected tasks. They should work together coherently.

Each recommendation unit should therefore contain an internal results chain, borrowing from results-based management (RBM). In simple terms, the chain distinguishes between activity — the discrete action undertaken; process — what actors do or put in motion; output — the concrete arrangement, product or capability that this creates; and outcome — the resulting change in institutional or professional practice. The purpose is not to force every recommendation into a rigid hierarchy, but to make explicit how the proposed actions are expected to lead to the change sought.



Method Note

The purpose of the results chain is to put order around the theory of change, not to suggest that reform unfolds neatly from one box to the next. Institutional change is continuous, uneven and often recursive, and organizations use terms such as activity, process, output and outcome somewhat differently. The practitioner should therefore focus less on perfect classification and more on the sequence of change: what does this action create or alter, and what does that in turn make possible?

These levels of change are not an abstract hierarchy of importance; their value depends on the diagnosis, the maturity of the reform issue, the level of institutional ownership, and the evidence available.

The main recommendation should sit at the level that corresponds to the problem and desired change. In some cases, that will be a change in institutional performance or professional practice. In others, especially where institutional commitment is weak, the evidence is still emerging, or the reform pathway is not yet operationally clear, the main recommendation may sit at output or process level.

For **example**, in a mature reform environment, the main recommendation may be:

Strengthen the consistent, lawful consideration and adjudication of confiscation in serious corruption cases where the legal and factual basis is present.

That targets changed practice.

In a less mature environment, the main recommendation may need to be:

Develop and adopt a confiscation-planning protocol for serious corruption cases.

That does not yet claim that practice has changed. It creates one of the conditions for practice change.

In an even earlier or lower-trust environment, the main recommendation may need to be:

Establish a structured institutional process to identify obstacles to confiscation planning in serious corruption cases.

That is a process-level recommendation. It may be appropriate where the first reform need is ownership, diagnosis-sharing, feasibility clarification or agreement on the route forward.

The point is to make the package coherent. The main recommendation should not become so broad that no one can act on it. The sub-recommendations should not become a list of activities with no visible connection to the intended change. Each sub-recommendation should answer a design question:

What condition does this element create, and how does that condition support the desired practice change?

A useful unit-level design test is:

If all sub-recommendations were implemented and used as intended, would the main recommendation become materially more likely to happen in practice?

If the answer is no, the unit is poorly structured. Either the main recommendation is too ambitious, the sub-recommendations do not address the real performance condition, or the results chain has not been made explicit.

Example

Starting formulation: Authorities should improve the use of confiscation in serious corruption and organized crime cases.

This is directionally right, but too broad. It does not identify where the confiscation chain breaks or which conditions must change for confiscation to be pursued more consistently.

A stronger recommendation unit would look like this:

Main recommendation:

Prosecution services, in co-operation with relevant investigative and financial-intelligence bodies, should institutionalize early confiscation planning in serious and complex OCC cases so that illicit gain, relevant assets and evidentiary needs are identified and developed as an integral part of the case strategy.

Possible supporting sub-recommendations:

- i. Prosecution services should introduce criteria for identifying cases that require early confiscation planning and require a documented asset-recovery plan for qualifying cases.
- ii. Investigative bodies and prosecution services should establish a structured handover through which available financial evidence, asset-tracing results and outstanding evidentiary needs are carried into prosecutorial case planning.
- iii. Prosecution services should ensure timely access to specialist financial-analysis support and introduce review points at which the confiscation component of the case is assessed before indictment and other significant procedural decisions.

The main recommendation states the desired practice change: confiscation should be planned from the outset rather than considered only at the end of the proceedings. The sub-recommendations create the enabling conditions identified in the theory of change: clear triggers and ownership, continuity of financial information and access to specialist support. Together, they form one coherent recommendation unit addressing early confiscation planning.

5 Assemble and test the reform recommendation package

Once the practitioner has drafted the units, they assemble them into the reform recommendation package. The package is not simply all recommendations placed under one heading. It is the operational form of the package-level theory of change.

The package should show:

- Which branch of the theory of change each unit puts into practice;
- Which units can proceed independently and which depend on another unit or enabling condition;
- How responsibility is distributed across institutions and whether the aggregate burden is realistic;
- How sequencing affects political support, resources, safeguards and the probability of uptake;
- Whether units overlap, contradict or create incentives that undermine one another;
- Which assumptions and evidence gaps apply to the package as a whole; and
- How implementation of each unit can later be followed without mistaking outputs for changed practice.

A package may combine, for example, one unit on ownership and workflow, another on professional tools and support, and a third on feedback and performance incentives. Each unit has its own main and sub-recommendations. Together, they address the combination of conditions identified in the theory of change. Where only one coherent unit is required, that unit may constitute the whole package.

Package coherence is therefore tested twice: internally, within every recommendation unit; and horizontally, across all units. The practitioner should be able to walk from each unit back to the diagnosis and forward to the milestones and verification evidence that Track will later develop.

6 Choose the adoption or disposition route and hand off to Track

Once the recommendation package is coherent, the practitioner must decide how it can realistically enter the reform environment. A technically correct recommendation that enters through the wrong route may be ignored, resisted, absorbed as a formal output, or pushed into an institution before the necessary conditions exist.

Some recommendations are ready for immediate institutional adoption. Others require staging, piloting, partner support, confidential technical dialogue, regional peer learning, or broader political alignment.

The adoption or disposition route should align with the evidence, lawful authority and the reforming institutions' implementation preparedness.

 Adoption or disposition route	 Use where	 What it means in practice
Direct recommendation	The responsible authority is clear, the change is within mandate, the evidence is strong, and action is feasible without major political or resource preconditions.	The recommendation can enter the consolidated reform package as a direct ask. This is appropriate for clear-cut priorities where the institution can act through an existing mandate, plan, procedure or reform commitment.


Staged recommendation

The final change is valid, but too large, sensitive or underprepared for one Step.

Retain the final practice or institutional change as the substantive target, but sequence implementation through one or more enabling conditions: clarify ownership, create a protocol, secure resources, establish a review routine, test a tool, or agree a sequencing plan. This is controlled sequencing.

Pilot or test route

Feasibility, institutional tolerance, access or operational viability is uncertain.

The recommendation should test a bottleneck before full roll-out. A pilot is useful only if it tests whether the proposed routine can work in practice; it should not become a small project detached from the reform logic.

Partner-supported route

The responsible institution has mandate and willingness, but lacks resources, expertise or implementation capacity.

The recommendation can be linked to a legitimate support vehicle, such as another international partner programme. The external partner supports the vehicle; it does not replace institutional ownership.

Regional peer-learning route

The issue is shared across jurisdictions, or a promising practice exists elsewhere and may be transferable.

The recommendation can be supported through regional exchanges, shared-issue tables, recommendation registries, deep dives or peer review. Transfer is never copy-paste; adaptation to the receiving legal and institutional context is part of the route.

Further-diagnosis route

A material uncertainty about the explanation, legal route, affected risks or feasibility prevents responsible design.

Specify the question, evidence required, responsible practitioner or institutional owner, decision value and stopping rule. Do not present further study as a substitute for a clear existing duty.



**Adoption or
disposition route**



Use where



What it means in practice

**Policy-ecosystem gap
flag**

The finding is real, but no current authority, mandate or vehicle exists.

The item should not be forced into a conventional recommendation. It may become an agenda item for a council, donor platform, steering mechanism, regional discussion or future diagnostic exercise.

**Deferred or no-
immediate
recommendation**

Action would overclaim the evidence, duplicate an existing process, create disproportionate risk or have little present prospect of materially affecting the problem.

Record the reason, preservation or referral action, review trigger and conditions under which the issue should be reconsidered.

This typology is not a ranking, and a direct recommendation is not inherently stronger than a staged one. A staged route may be more methodologically appropriate where it accurately identifies the enabling condition that must be met before the substantive target can be pursued or sustained.

As with the development and selection of the recommendation itself, adoption or disposition routes are decisions based on the best available evidence, not permanent classifications. Track may later show that a direct recommendation needs staging, that a staged recommendation is ready for implementation, that a pilot should be modified, stopped or scaled, that partner support is distorting ownership, or that an ecosystem gap has become actionable.

► Before you move on



ETHICAL FRAME NOTE

Recommend is where the engine proposes institutions to act, and the binding constraint is judicial independence and prosecutorial autonomy. No recommendation may set targets for conviction, sanction, speed or confiscation, direct how a case should be decided, or use performance targets in a way that pressures decision-making or trades away fairness, reasoning or lawful discretion. A measure touching case assignment, workload or review must preserve decisional independence; where it could blur that line, narrow or redesign it before it leaves this Step. Affected institutions, professional groups and, where appropriate, rights-holders should have a proportionate opportunity to test feasibility and likely effects before a recommendation is final. Check whether burdens fall unevenly on smaller or under-resourced offices, and whether foreseeable harm — retaliation, reputational damage, resource strain, perverse incentives — can be mitigated. Where it cannot, narrow, stage or reconsider.



COMMUNICATION CARD

At this Step, communication is used to pressure-test how a proposed practice change can work inside the real institutional system. Bring the reform problem and the desired change into the conversation clearly enough that counterparts can engage with the mechanism, rather than simply reacting to a recommendation sentence: “this is the practice we are trying to change; this is what the evidence suggests is preventing that; what route could realistically change it?” Counterparts may show that a proposed implementation lead lacks authority or the proposed addressee is not competent, that a mechanism already exists, or that sequencing is unrealistic, all useful inputs to recommendation design. What should not move is the underlying diagnosis. Unless new evidence genuinely reopens Analyse or Explain, the finding and the practice change it points to are not negotiable because the intervention is inconvenient. Ownership comes from participation in designing the solution, not from a veto over the finding.



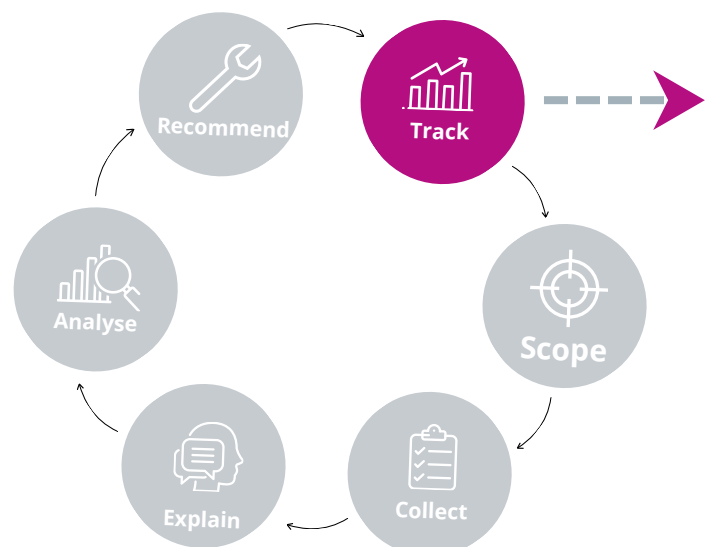
ADAPTATION NOTE

The method remains consistent across reform fields: preserve the diagnosis-to-option chain; confirm the law, authority, rights and safeguards; generate alternatives; state the expected contribution; and design for later verification. What changes is the applicable law, the institutional chain, affected groups, confidentiality risk, political economy, resources and the feasible evidence base. A ministry or legislature requires attention to legislative competence, constitutional and regulatory review, budgetary effects and implementation responsibility. Judicial or prosecutorial governance bodies require specific safeguards for independence and autonomy. An individual court or prosecution office may need a bounded operational recommendation and an escalation route when the relevant condition falls outside its authority. Civil society and international actors must distinguish between independent accountability, advice, convening, funding and implementation support, and must document positionality and domestic ownership. Whatever the setting, the evidence standards, legal safeguards and handover to Track set out in this Step should remain intact; what adapts is the applicable law, the actors involved, and the surrounding reform context.



STEP-CHECKLIST

- ☐ Is the **theory of change traceable to the finding and explanatory account**, and does it identify the conditions that must shift for practice to change?
- ☐ Does each **recommendation unit** contain a clear main recommendation and supporting sub-recommendations that form a **coherent results chain**?
- ☐ Do the units collectively put the **package-level theory of change into practice**, without material gaps, duplication or contradiction?
- ☐ Is every **proposed action assigned to a lawfully and realistically positioned addressee**, decision authority or implementation lead, as applicable, together with an appropriate adoption or disposition route?
- ☐ Have the **selected measures been tested for technical soundness, political supportability and operational feasibility**, with relevant assumptions, risks and trade-offs recorded?
- ☐ Is the **package formulated clearly enough for Track to identify implementation milestones** and assess whether the expected conditions and practices are beginning to change?



Step 6

► Where you are in the cycle

Track receives a reform recommendation package and, for each recommendation unit, traces movement from institutional Action, through Uptake in ordinary work, to observable Change and, once Change is established, whether it is sufficiently durable to be regarded as institutionalized. Reform rarely follows this pathway neatly or through the exact route originally recommended, so Track also asks whether the evidence confirms the anticipated pathway, reveals a bottleneck or divergence, or shows that change is occurring through another route. Track does not end the policy cycle; it routes learning onward, whether by continuing Track, returning to Explain or Recommend, or handing a defined evidence need back to Scope or Collect.



► Why this Step is important

Tracking keeps the reform dialogue alive and offers a structured way to recognize progress, identify bottlenecks, adjust expectations and distinguish formal compliance from actual change in practice. This is where the Guide returns to its central EBTM principle: effectiveness is not established by institutional assertion or formal activity alone; it must be demonstrated in observable practice. A new law, strategy, guideline or training programme may be important evidence of Action, but it does not by itself establish Uptake or Change.

► What happens if we are not intentional about it

Without tracking, recommendations remain well-written advice. They may be welcomed, resisted, partially implemented, stalled, or replaced by a formal routine that only appears to address them. We would not know which occurred, why, or whether the original explanatory account still holds. Conversely, a reform may be dismissed as unsuccessful because it has not yet produced full institutional change, even though the necessary groundwork has been laid.

► What you produce

By the end of this Step, the practitioner should have:

an evidence-based account of Action, Uptake and Change for each recommendation unit, supported by predefined implementation milestones and a documented evidence trail; an implementation narrative explaining what has moved, what remains blocked, how the enabling conditions are functioning, whether the reform is following the pathway anticipated in Recommend or a different route, and the assessment's level of confidence; and a clear next-action decision and route back into the engine, identifying the Step best placed to act on it.



THE STEP AT A GLANCE

1 Set the tracking framework methodology

What would meaningful movement actually look like?

Build the methodology from its two connected tools: the three reform questions – Action, Uptake and Change – which keep different kinds of movement distinct, and the implementation milestones that state in advance what each kind of movement would look like.

2 Set the tracking matrix

What do we record and review for an objectively verifiable assessment?

Build the recommendation unit's minimum monitoring plan: what is being tracked, the tracking baseline, the material milestones, which part of the reform pathway each one informs, the evidence sources, the relevant actors and dependencies, and the review period. Defining these before evidence is collected is what reduces hindsight bias and preserves the link from theory of change to evidence to judgement.

3 Collect and verify evidence of reform implementation

What kind of reform implementation movement has occurred?

Return where possible to the evidence base that supported the original finding, adding new sources where implementation creates new evidentiary needs – including a targeted return to Scope where a milestone requires a fresh cohort of cases. Record source, date, scope and limitation for each item, treat institutional accounts as institutional evidence rather than verification, keep material variation visible, and separate evidence of action from evidence of practice.

4 Interpret the data on reform implementation progress

How far has the reform travelled, and what does that mean?

Read Action, Uptake and Change together at the level of the recommendation unit. Write the implementation narrative, state confidence and limits, and record a question as not assessable where the evidence cannot answer it.

5 Route implementation learning back into the engine

What has this taught us, and which question should the engine answer next?

Choose one route and record it: back to Scope, Collect, Analyse, Explain or Recommend; continue tracking; or close a recommendation.

► What Track does: tracking implementation begins with the recommendation package

Track is the periodic, evidence-based assessment of whether a recommendation has been acted upon, adopted in everyday work, and followed by the intended change in practice. It is not an audit, a compliance inspection, an institutional ranking, or a full impact evaluation. Its purpose is to make implementation visible enough for reform actors to recognize progress, identify where the pathway is breaking down, and decide what to do next.

A reform recommendation package typically comprises several recommendation units, each with a main recommendation that captures the central change sought, aligned with the evidence, mandate, and scope, and related sub-recommendations that provide the enabling Steps to make that change possible. Evidence and milestones for sub-recommendations are tracked separately where useful, but sub-recommendations do not receive stand-alone implementation judgements; the assessment is synthesized at the recommendation-unit level, with each sub-recommendation treated according to the enabling function it was designed to perform. For example, a package addressing delayed confiscation decisions might combine a main recommendation that decisions on confiscation be issued, where the matter is properly before the court, within the applicable statutory timeframe and with the reasoning required by law, with sub-recommendations assigning case-management responsibility to a designated unit, adopting a standard financial-evidence checklist, and providing targeted training for prosecutors handling asset-recovery cases.

Track follows each unit over time; different units within the same package may progress at different rates and along different pathways. Within a single package, one unit's main recommendation may be progressing steadily while a related sub-recommendation stalls or proves unworkable in practice. Track should first ask whether the function that sub-recommendation was meant to perform remains necessary or is being performed through another route. The unit returns to Recommend only where the gap materially undermines the design or implementation pathway; otherwise Track records the divergence and continues to assess the unit.

One of the most common errors in reform monitoring is to ask only whether a recommendation is completed. Between a recommendation being accepted and practice genuinely changing lies a long middle passage: institutions recognize the problem, assign ownership, develop rules and tools, begin using them, and eventually incorporate them into ordinary work. In this middle passage, recommendations either become symbolic or genuinely change practice. Track is both follow-up and learning; it assesses implementation and tests the recommendation's design, the explanatory account, and the evidence base linking the observed problem to the recommended response.

1 Set the tracking framework methodology

Track concerns change over time, not merely whether a recommended action was completed, but whether the reform has entered the institutional system, begun to be used, and altered the practice or condition it was intended to affect. A useful methodology must preserve the distinctions between these forms of movement rather than collapsing them into a binary determination of implementation.

The methodology rests on two connected tools, reform questions and implementation milestones. Three linked reform questions – asking about Action, Uptake and Change – organize both the design of those milestones and, later, the interpretation of the evidence gathered against them. Implementation milestones define, in advance, the evidence-based markers of meaningful progress along the reform pathway. Neither tool removes professional judgement nor creates a mechanical measure of reform; they structure judgement.

1.1 Reform questions assessing Action, Uptake and Change

Track applies three reform questions:

1. What **Action** has been taken?
2. Is there an **Uptake** of relevant measures and practices?
3. What relevant **Change** is evident?

Each of the three questions describes a different kind of movement. Within Action, Uptake and Change, it is critical to understand the shape of that movement, described using implementation stage descriptors, because showing some progress and demonstrating actual change in practice are not the same claim. The descriptors below are grouped under Action, Uptake and Change and show what progress on each question can look like. They are not a fourth taxonomy or a single seven-stage implementation ladder, and they should not be collapsed into one overall stage. A portfolio overview may report the question-level descriptors side by side, but the underlying Action, Uptake and Change answers and the narrative remain the substantive assessment.

1.1.1 Describing progress on Action

Action asks whether a formal commitment has been made and whether Steps have been taken to advance the recommendation within the reform process. Evidence may include ownership, planning, assigned responsibilities, resources, consultations, adopted rules, training, tools, workflows, or institutional decisions. Do not treat activity counts alone as sufficient evidence of meaningful implementation.

Action uses three recognizable implementation stage descriptors:

- **No Verified Action:** The available evidence supports no verified entry of the recommendation into the reform process. If the evidence is unavailable or inadequate to answer the question, record Action as not assessable instead of assigning this label.
- **Commitment:** Those responsible recognize the need for change, and credible intent or ownership is evident, but concrete preparation has not yet begun.
- **Planning:** Concrete preparation or formal action has begun; responsibilities, plans, rules, resources or tools are being assembled, adopted or established, but use in ordinary practice has not yet begun.

For example, a ministry publicly acknowledging a delay problem and naming a responsible unit shows Commitment; that unit circulating a draft workflow and a resourcing request shows Planning. An adopted rule, completed tool or other formal measure that has not yet entered ordinary use also remains at Planning for the Action question. None of these, on their own, shows that practice has changed; the next two questions assess use and Change.

1.1.2 Describing progress on Uptake

Uptake assesses whether resulting measures are used and applied in institutional or professional practice after the formal adoption, which belongs under Action. Evidence should demonstrate repeated application, participation by the intended actors, operation of the relevant workflow, or references in judicial or administrative acts. Uptake is often uneven across institutions, case types, professional groups, or leadership periods. Variation is important evidence, so keep it visible rather than collapsing it into a single conclusion.

Uptake is most closely associated with:

Early Implementation: Material measures have moved beyond preparation and entered actual use, but the Change carried by the recommendation target is not yet established.

The depth of Uptake may nevertheless vary considerably within this stage. Initial or partial use and routine, consistent use are analytically different, even though both remain within Early Implementation until the evidence demonstrates Change. For example, a small number of prosecutors piloting a new financial-evidence checklist indicates early Uptake; the checklist becoming a routine Step across the designated unit indicates much stronger Uptake. Neither, by itself, establishes Operationalization unless the practice or institutional condition that the checklist was intended to change is observably different.



Method Note

The implementation stages describe **change along a continuum**, not six discrete states into which reform fits perfectly. Movement can therefore occur within a stage before the defining condition of the next stage is reached.

This is particularly important for **Early Implementation**. Uptake may range from initial or uneven use to regular and widespread application of a new arrangement. Both can remain within Early Implementation where the central Change addressed by the recommendation has not yet been established. The difference should be captured in the evidence and implementation narrative rather than by creating infinite stages.

The stage changes only when the **defining threshold of the next stage** is supported by the evidence. Routine use alone does not establish Operationalization; Operationalization begins when the relevant practice, process, output or institutional condition is observably different in ordinary work.

The same principle applies throughout the ladder: the stages provide common reference points along a continuous reform pathway, while the narrative captures the movement, variation and distance between them.

1.1.3 Describing progress on Change

Change asks whether the practice, output or process specified in the main recommendation is observably different or functioning as intended. This is the critical EBTM threshold: activity and uptake matter, but assessment must be made at the level of practice that the recommendation sets.

Change uses two recognizable implementation stage descriptors:

Operationalization: The intended practice change is evident and reasonably consistent in everyday work.

Institutionalization: As the descriptor of higher and more durable change, the practice is sufficiently sustained and supported by ordinary institutional conditions to be considered durable, without depending on individuals or temporary or exceptional support.

For example, where confiscation is properly before the court, consistently issuing a reasoned decision within the applicable statutory time frame may demonstrate Operationalization. The Institutionalization test then asks whether that continues despite changing factors, such as shifts in personnel or funding.

Taken together, these six descriptors provide a common vocabulary for interpreting and reporting movement along the reform pathway. They allow practitioners to communicate progress consistently across recommendation units and review periods while preserving the nuance captured through the three Track questions, milestone evidence and implementation narrative. The stage gives the reader a concise reference point; the narrative explains where within that stage the reform sits, what is moving, what remains uneven or blocked, and what would justify progression to the next stage.

1.2 Implementation milestones

An implementation milestone is a predefined, evidence-based marker of meaningful progress. It indicates what the practitioner would expect to see if a relevant part of the recommendation unit were beginning to perform the function assigned to it under the theory of change. It sits between the conceptual question, “What do Action, Uptake or Change look like here?” and the later evidentiary question, “What do the available sources show?”

There is no fixed number of milestones a recommendation unit must have. As a practical guide, most units need at least one milestone for each of Action, Uptake and Change. A unit with milestones only at the Action level cannot support a Change judgement, however many Action milestones it has. Beyond that minimum, add milestones only where they resolve a genuine uncertainty about the pathway, not to make the matrix appear more thorough than the evidence base can support.

1.2.1 Aligning the milestones with Action, Uptake and Change

Milestones should align with the recommendation unit and with the function each component performs. Action and Uptake milestones will often relate to the enabling sub-recommendations, while Change milestones will often relate to the main recommendation, which carries the central recommendation target. This allocation is not fixed, and the practitioner should classify each milestone according to the evidence it demonstrates.

Different components of a unit perform distinct functions and need not follow the same sequence. A sub-recommendation assigning institutional ownership may primarily need Action milestones; a new protocol may need both Action and Uptake milestones; a recommendation concerning professional practice may need Change milestones directly.

This reflects the typical reform progression, in which ownership and design measures tend to move first, use of the resulting tools and procedures follows, and change in the underlying practice follows that. However, the progression is neither mechanical nor linear. Reform rarely moves neatly from one level to the next: a measure may be adopted and used quickly in one unit while stalling entirely in another, or practice may shift before every originally recommended supporting measure is in place. Track should record the sequence it observes rather than the one it expected.

Durability is not a separate milestone level that every sub-recommendation must reach independently. In assessing Change at the recommendation unit level, milestones on individual sub-recommendations, such as stable ownership, recurrent resourcing, and established feedback arrangements, serve as evidence for the unit-level Institutionalization test, rather than as separate durability judgements. This distinction keeps the methodology aligned with Recommend: a sub-recommendation enables the main recommendation, not by accumulating its own progress markers.

For example, a recommendation unit tasked with improving the publication of judgements might set the following milestones:

Action: A publication protocol assigning responsibilities and setting the anonymization workflow is formally adopted.

Uptake: Responsible staff and courts apply the protocol when processing relevant judgements.

Change: Relevant judgements are published within the applicable timeframe, with the required anonymization and accessible reasoning.

Another reform recommendation package to strengthen prosecutorial handling of financial evidence in asset-recovery cases might instead include the following milestones:

Action: A specialist financial-evidence support and training programme is formally established.

Uptake: Trained prosecutors use the financial-evidence checklist when preparing indictments in monitored cases.

Change: In monitored cases alleging illicit gain, the confiscation component of the indictment meets a predefined, legally grounded content criterion, for example by identifying the alleged gain, the evidentiary basis relied upon where required, and the legal basis for any requested measure.

The milestones in the two examples place different demands on the evidence base. The first relies heavily on case-processing records and court publication logs, whereas the second relies on indictment content across monitored cases and the records needed to verify when relevant financial evidence became available. This underscores the importance of defining milestones and their evidence sources together, rather than treating milestone design and evidence collection as separate stages.

1.2.2 Articulating the milestones

A good milestone translates part of the expected reform pathway into an observable proposition that is specific enough for different practitioners to recognize the same phenomenon. Whether quantitative or qualitative, it must clearly state what should be observable, for whom, within what scope, and in what direction. Four disciplines help:

State what changes, not what improves. Evaluative adjectives cannot be observed, only asserted.

Weak: Reasoning quality in confiscation cases improves.

Stronger: Judgements in monitored cases in which confiscation is properly before the court address illicit gain and provide the reasoning required by the applicable legal standard.

The first formulation requires the practitioner to import a personal standard for what counts as an improvement; the second can be checked by comparing the judgement with the indictment.

One coherent observable proposition per milestone. Do not combine independent stages or conditions that could move separately, such as adoption and use.

Weak: Protocol adopted and applied in practice.

Stronger: Separate into two milestones: first, the protocol is adopted, and then the protocol is applied in practice.

This also keeps adoption (an Action milestone) analytically distinct from application (an Uptake milestone).

State the direction and, where one exists, a reference point.

Weak: Relevant decisions are published more quickly than in previous years.

Stronger: An increase in the proportion of relevant judicial decisions published within the legally prescribed timeframe, with the tracking baseline established for the recommendation unit.

The reference point, measured against the tracking baseline, need not be numerical; it may be derived from an applicable legal standard, baseline practice, or the defined case population. Use numerical thresholds only when they have a defensible basis, not merely to appear measurable; otherwise, they risk becoming arbitrary targets the institution can appear to meet without resolving the underlying problem.

The same discipline applies to timing. Where timing is intrinsic to the reform itself, for example, meeting a statutory deadline, the milestone should state the applicable timeframe directly rather than leaving it implicit. Where timing is not intrinsic to the substance of the milestone, avoid attaching an arbitrary date merely to create a sense of measurability; instead, state when the next review will assess it.

1.2.3 Using existing metrics

Where possible, milestones should draw on metrics already embedded in domestic strategies, action plans, or institutional monitoring systems to strengthen ownership and reduce parallel reporting burdens. Adoption should not be automatic. Test each metric against the recommendation logic to determine whether it demonstrates progress towards the intended practice change or merely indicates that an activity occurred. Consult Scope for any gaps already identified in these data sources.

Official metrics often count outputs, such as training sessions delivered, rules adopted, and cases registered, because outputs are easier to measure than practice. An output count can still be a legitimate Action or Uptake milestone, but it should not be relabelled as evidence of Change without practice-based or otherwise claim-fit evidence that the counted activity altered how cases are handled. Where the official metric falls short of the practice-level question the recommendation raises, supplement it with a milestone that Track defines directly, and note the gap so anyone relying on the official figure alone can see the limitation.

1.2.4 Reliability, feasibility and utility filters

A milestone remains provisional until it has been assessed for reliability, feasibility, and utility.

Reliability: Can the milestone be interpreted consistently?

Weak: Greater transparency is visible in judicial practice.

Stronger: Relevant judicial decisions are published within the legally prescribed timeframe.

The first leaves “greater” and “transparency” open to interpretation; the second can be assessed against a fixed timeframe and a defined population of decisions.

Feasibility: Can the milestone realistically be followed?

Weak: Financial intelligence relevant to confiscation is provided to prosecutors sufficiently early to inform charging decisions.

Stronger: In monitored cases where the prosecution file records receipt dates, the financial evidence cited in the indictment was recorded as received before the charging decision.

The first may be highly relevant to the theory of change, but Track often cannot verify what internal information a prosecutor had at a given moment. Feasibility should not be an excuse for selecting only what is easy to observe. If an important part of the theory of change cannot be observed directly, state this clearly.

Utility: Does the milestone tell us something that matters for the reform judgement?

Low utility: Number of press releases announcing the transparency reform.

Higher utility: Relevant judgements are published in accessible form within the applicable time frame.

The first may demonstrate communication activity but tells the practitioner little about whether transparency has improved for those who need the decisions.

A milestone can be reliable and easy to collect while still adding little value. Ten reliable Action milestones do not compensate for the absence of a way to examine Uptake or Change. Consider utility across the milestone set as a whole, not only milestone by milestone: a set that speaks only to Action, however precisely worded, leaves the central practice-level question unanswered. Where an important part of the theory of change cannot presently be followed, record that as an evidence gap rather than substituting a weaker but more convenient milestone.

The practical test for milestones:

- Can we interpret this consistently?
- Can we realistically follow it?
- If we observe it, will it change what we can responsibly say about reform progress?

1.2.5 When milestones reveal a design problem

Tracking also tests the recommendation itself. If no meaningful milestone can be set, the recommendation may be too vague, too ambitious, directed at an actor without the necessary mandate, or disconnected from the diagnosed condition; the problem is design, not measurement, and the recommendation may need rewriting in Recommend.

For example, a recommendation that a ministry “strengthen inter-agency co-ordination” resists milestone-setting because it names no observable action, no responsible actor, and no defined practice to change; that difficulty is itself useful information and should be routed back to Recommend as a design issue rather than absorbed by inventing a milestone the recommendation was never specific enough to support.

2 Setting the tracking matrix

Before assessing progress, the practitioner should build a tracking matrix which functions as the recommendation unit’s minimum monitoring plan. It should make clear what is being tracked, the baseline position, the material milestones, which part of the reform pathway each milestone informs, the evidence sources, relevant actors or dependencies, and the review period.

Its value is methodological as much as administrative. By defining these elements in advance, the practitioner reduces hindsight bias, keeps successive reviews comparable and preserves the link between the theory of change, the evidence collected and the eventual maturity judgement.

The matrix should remain proportionate to the recommendation. It is not a catalogue of every activity relating to the milestones. Include only the signals needed to understand whether the material reform pathway is moving and where it may be blocked. Individual rows track evidence; they are not scored or assessed separately. The final judgement is made at the level of the recommendation unit as a whole.

3 Collecting and verifying evidence of reform implementation

3.1 Evidence sources

Wherever possible, Track should return to the same evidence base that underpinned the original finding and recommendation, including hearing observations, case reports, legal acts, procedural histories, official statistics, access requests, media and contextual monitoring, consultations, End-of-case assessments, and Deep Dives. This preserves methodological continuity in the reform inquiry and helps prevent follow-up from drifting into a parallel exercise that assesses implementation solely through institutions’ own accounts of their activity.

Consultations warrant caution. A senior official’s account of what has changed is a legitimate and often necessary source, particularly for understanding Action, where institutional intent is otherwise invisible. However, it is an institutional source like any other and should be weighed accordingly, not treated as direct evidence of practice. Where a consultation is the only available source for a Change milestone, record that limitation explicitly in the confidence statement rather than allowing an institutional account to silently substitute for practice-based or otherwise claim-fit verification.

This does not restrict Track to the sources originally used to diagnose the problem. Implementation creates new evidentiary needs, and new sources may be added where relevant and proportionate, provided their limitations are recorded. For example, if the original recommendation concerned the quality of reasoning in confiscation judgements, Track should consult the judgements themselves rather than relying solely on the court’s case-management statistics, which may show that a case was closed without indicating how it was reasoned.

Implementation tracking may also require new practice evidence generated after the reform. If a milestone requires a new cohort of cases, Track may reopen Scope to define a targeted new monitored pool. Tier 2 monitoring, a modality primarily reliant on written sources, can often provide a proportionate way to test a specific implementation question without re-establishing full longitudinal monitoring of every case. Define the selection boundary, observation period, and monitoring depth explicitly to ensure the new evidence remains comparable to the tracking baseline.

3.2 Gather and verify

Tracking should be periodic and regular; one assessment captures a moment, while repeated assessments capture change over time. Set a review frequency appropriate to how quickly the relevant practice can plausibly change, and record the source, date, scope, relevant item and milestone for each piece of evidence, along with any limitation affecting interpretation. Collect proportionately and cross-check material claims. Independent corroboration is stronger than repeated versions of the same institutional assertion; where the institution responsible for implementation provides the evidence, treat it as institutional evidence rather than as independent verification.

When evidence is absent or inaccessible, distinguish among no verified action, unavailable evidence, and a genuine evidence gap; absence of verification is not evidence of non-occurrence. Assess whether Action, Uptake or Change is consistent across institutions, case types, professional groups or time periods, and keep material variation visible rather than letting it disappear into a single conclusion. Recent measures may need time before evidence of practice change emerges, and implementation may also regress, so account for timing and plan for periodic review.

Access to evidence is not always straightforward. An institution may respond slowly, provide only aggregated figures when case-level detail is required, or decline to share internal documents. When this happens, record the request, the response, and the date, rather than leaving the gap unexplained. A documented access limitation is relevant, and a pattern of limited access across several reviews may warrant assessment.



Practitioner Note

A brief log kept alongside the tracking matrix is usually sufficient to support the judgement and explain why the practitioner reached a particular conclusion. It need not be elaborate; the discipline lies in keeping the record contemporaneous, not in its format. This log should capture each piece of evidence, recording the source, date, the milestone it addresses and its evidentiary role - for example, institutional self-report or a record of practice - and should note independent or meaningfully complementary corroboration where relevant to the claim.

3.3 Distinguish evidence of action from evidence of practice

This distinction matters because the two kinds of evidence are easily conflated and often reside in different places: action evidence is usually produced and held by the institution itself, whereas practice evidence typically comes from the cases, acts and administrative use through which the relevant practice becomes visible, rather than from implementation self-report alone.

Workplans, budgets, minutes and training records may show an intent to change practice; legal acts, case trajectories, hearing observations and administrative use show whether practice has changed.

These evidence categories track directly onto the three questions:



Action

Statements, strategies, workplans, decisions, minutes, budgets, staffing decisions, adopted rules, training records, systems created.



Uptake

Use of the new tool or workflow, participation by intended actors, repeated application, references in judicial acts, publication practice.



Change

Recommendation target is observably different, including as demonstrated through changed case trajectories, reasoning patterns, timeliness, access experience, routine compliance, sustained use across time or units.

4 Interpreting the data on reform implementation progress

By this point, the practitioner has defined milestones and evidence sources and gathered and verified the available implementation evidence. The task now is to determine what that evidence establishes about the recommendation unit.

4.1 Read the three questions together for the recommendation unit

Recall that the three questions serve as a framing device for the recommendation unit, not as three separate assessments completed independently for each component. The main recommendation sets the target; the sub-recommendations define the conditions expected to enable or sustain that change.

Extensive Action may coexist with weak Uptake. New arrangements may be implemented without achieving the expected Change. Change may emerge even when some originally recommended supporting measures remain incomplete. This divergence is not a defect in the methodology; it is information about how the reform is progressing and may prompt the practitioner to revisit a dependency, the theory of change, or the recommendation itself.

In practice, most recommendation units will show a mix of evidence and question-level statuses across their components at any given review. Reporting this mix, rather than collapsing it into a single overall assessment, is the point of reading the three questions together, as this shows where the pathway is blocked, not merely how far the unit has progressed.

4.2 Interpretive discipline

Three disciplines guide the interpretation used to answer the three questions:

Read function, not volume: Recommend selected each sub-recommendation because it performs a material function in the theory of change. Ask what the movement of those conditions means, not how many measures have been completed. A blocked gateway condition does not disappear because several more visible activities have advanced.

Distinguish reform machinery from the recommendation target: Action and Uptake can carry a unit through substantial implementation, but no quantity of plans, procedures or tools can substitute for evidence that the practice or condition carried by the main recommendation is functioning differently in ordinary work.

Do not make the original pathway a compliance test: Recognize when Change is verified through an alternative pathway, namely measures or conditions that differ from those originally anticipated in Recommend. Ask what conditions may have contributed to and now support or sustain the Change, not whether the institution followed the recommended sequence. Missing supporting measures remain relevant to the durability test but do not automatically invalidate verified Change.



Practitioner Note

Be aware of two potential errors.

The first is scoring by volume: Treating a long list of completed Action milestones as evidence of Uptake or Change, even though a single blocked gateway condition can prevent movement along the expected pathway even when other Action milestones are complete; the list should not be converted into a single overall status assessment.

The second is anchoring on the original plan: Continuing to look for evidence along the exact pathway Recommend anticipated, and reporting “not implemented” when the evidence instead shows Change arriving by a different route, or a durability question that the original plan never addressed.

Both errors are easiest to avoid by returning to the specific evidence for the milestone, rather than forming an overall impression first and then selecting evidence to support it.

4.3 Explain the assessment through the implementation narrative

The three-question assessment condenses the judgement into a concise account; the narrative provides the necessary detail and meaning. It should explain what moved and what remained stalled; which milestones and evidence were decisive; whether movement remains at the level of ownership and preparation or has reached ordinary practice; which enabling conditions matter most and whether any material bottleneck remains; whether the pathway anticipated in Recommend is holding or change is occurring through another route; whether the observed change appears broad, fragile, or institutionally supported; what contrary evidence or uncertainty remains; and what this means for the next review.

4.4 State confidence and limits

The narrative should make clear how strongly the evidence supports each material conclusion, identifying the principal supporting evidence, material contrary evidence, important variation, and the limitation most likely to change the judgement at the next review. When the evidence is strong and consistent, state the conclusion directly.

When it is limited, mixed, recent, or confined to a narrow sample, use correspondingly cautious language, such as “early evidence indicates”, “the available evidence suggests”, or “there is not yet sufficient evidence to determine whether the change extends beyond the monitored scope”, so the reader understands what would justify a stronger or different conclusion next time.

When the available evidence is genuinely insufficient to answer Action, Uptake or Change with any confidence, record the question as not assessable rather than defaulting to the lowest status descriptor. Not assessable is not an implementation stages descriptor under Action, Uptake or Change; rather, it is a statement about the evidence base. It should specify what evidence would resolve the question and when it might become available. Treating an evidence gap as equivalent to “no progress” risks penalizing a recommendation unit for a monitoring limitation rather than an implementation failure.

Examples

1.Reform machinery advances, but practice does not:

A co-ordination mechanism has been established, procedures adopted, and specialist training embedded, yet monitored practice shows no corresponding change in the way the law is applied. Action and Uptake are in place; Change is not. The volume and quality of supporting activity do not, by themselves, move the recommendation unit into Operationalization; no number of co-ordination meetings or training sessions substitutes for evidence that cases are being handled differently.

The next review should assess whether the gap reflects a reasonable time lag, an unresolved bottleneck, absorption into existing routines, or a weakness in the original theory of change. Uptake cannot compensate for the absence of Change.

2.Practice changes before the recommended structure is complete:

Monitored evidence shows that the application of the law has become more consistent, but some originally recommended supporting measures, namely a co-ordination body and a training programme, were never established.

Change is established (Operationalization) because the defining condition has been met, as the practice is visibly functioning differently in monitored cases, and an incomplete structure does not erase that. The next question is which conditions may have contributed to and now support or sustain the Change, since the reform may have taken a route Recommend did not anticipate. Informal peer learning, a change in leadership, or an unrelated procedural reform may be relevant explanations to test. The unit does not reach Institutionalization unless the evidence also shows that the changed practice is sufficiently supported by durable arrangements, whether the originally recommended ones or a substitute performing the same function, to endure.

5 Route implementation learning back into the engine

Track does not close the policy engine by declaring success or failure; it closes the loop by generating learning. Sometimes the evidence shows progress; sometimes it shows stalled implementation, substitution, or regression; sometimes a recommendation unit proves too broad, misdirected, insufficiently sequenced, or based on an incomplete diagnosis.

These are not failures of the engine, so the implementation narrative should ask what the evidence has taught and which question the engine now needs to answer.

 Route	 Use where
Return to Scope	A new implementation question requires a defined monitored pool, observation window, evidence strategy or monitoring depth.
Return to Collect	Relevant implementation evidence is missing, inaccessible or not yet organized into a usable evidence base.
Return to Analyse	New practice evidence reveals a recurring or structurally significant pattern or verified common rule, workflow or institutional arrangement that Analyse must bound and interpret.
Return to Explain	Implementation occurred but the expected change did not follow, or the main recommendation and sub-recommendations diverged in ways the original mechanism cannot explain.
Return to Recommend	The recommendation is untrackable, wrongly directed, infeasible, insufficiently sequenced, or no longer the best route to change.
Continue Track	The pathway remains plausible, but more time or repeated evidence is needed to assess uptake, routine use or durability.
Close or retire	The intended practice is sufficiently established for the purpose of the inquiry, the issue has materially changed or continued tracking no longer adds value.

When Track cannot determine whether a recommendation has advanced because the evidence is unavailable or insufficient, that is itself material information about the evidence environment. Route the specific need to Scope or Collect rather than treating the absence of verification as evidence of no action. A route need not be exclusive: an unexplained pattern may require a return to Explain while Scope simultaneously defines a new evidence stream. The practitioner should first identify the primary unresolved question.

Not every gap requires redesign. Reform takes time, and the absence of immediate change does not, by itself, mean the recommendation has failed. Continuing to track a plausible pathway is often the right call, provided the practitioner records what the next review is expected to establish rather than simply repeating the review indefinitely. For example, the recommendation unit remains at Established Usage; the next review will test whether continued use of the revised workflow has begun to shift the timeliness figures it was designed to affect.

Progress and contribution are distinct claims. If Change is demonstrable, recognize it without automatically crediting the recommendation. Track may assess whether the evidence supports a bounded contribution claim but should not treat temporal sequence or association as proof of causation. If a contribution claim is not supported, record the Change as verified, leave the cause open, and route the explanatory question to Explain. Similarly, “not implemented” is rarely a sufficient explanation on its own; the evidence may instead point to a specific blocked condition, a resourcing gap, or a design flaw that the practitioner can identify.

Closing a recommendation does not mean the wider reform problem is permanently solved; it only means that continued tracking of that specific recommendation no longer adds sufficient value within the purpose and scope of the inquiry. A recommendation unit may be closed where the intended change has become sufficiently established and durable for the purpose at hand, where the underlying issue has changed materially so that the original recommendation is no longer relevant, or where continued tracking would not alter the judgement or the response. Record the reason and retain the evidence trail; a later cycle may reopen the issue if practice regresses.

► Before you move on



ETHICAL FRAME NOTE

At Track, several principles are under pressure because this Step produces condensed assessments of institutional performance.

Objectivity requires reading each question against the available evidence, never inflating it to flatter a co-operative counterpart, never deflating it to satisfy an external appetite for candour, and never letting it become a target the institution learns to game. Milestones are markers used to assess movement towards reform, not compliance targets for the institution. Every conclusion carries a documented rationale, including the milestones, the evidence, and the reasoning behind them, so the judgement is explicit rather than hidden.

Confidentiality governs the disclosure boundary this Step must maintain: the underlying evidence file and full reasoning remain internal, and what reaches an external audience should be calibrated to the mandate, the confidentiality of sources, and the reporting product—not treated as an all-or-nothing choice.

Do not rank institutions or turn the assessment into a competition. Comparative information may be analytically necessary to explain variation, but it should not produce league tables. Track assesses an institutional or recurring practice within the defined scope; it does not grade individual judicial or prosecutorial decisions, and it must respect judicial independence, prosecutorial autonomy and the non-intervention principle, beyond what the evidence and mandate require.



COMMUNICATION CARD

At this Step, communication is used to make implementation movement intelligible while keeping activity, uptake, change and durability distinct. Begin with what the evidence establishes, then name the next evidence needed: adoption of the guidance completes an enabling milestone indicating early implementation; to assess operationalization, the next evidence should show whether the guidance is routinely used in relevant casework. This avoids both inflating formal activity into effect and dismissing genuine enabling progress as though it did not count. The most useful implementation story is usually the transition that has stalled. So, in turn, ask institutions for the evidence corresponding to the next meaningful transition guided by the milestones. If they report that a procedure is now routinely used, what record would make that use observable? If a new workflow exists, where should its effect appear in the original evidence spine? Be aware though that institutional self-report is evidence of what the institution reports was done, not that a practice is used, effective or sustained. Equally, do not use Track as a compliance contest. Its purpose is action-learning, not institutional grading- interact with the interlocutors accordingly.



ADAPTATION NOTE

The discipline remains the same. For example, a practitioner monitoring the implementation of a prison-conditions recommendation would use the same structure with different content: Action might be evidenced by a revised inspection protocol and a designated oversight unit; Uptake by inspectors applying the protocol during scheduled visits; Change by measurable improvement in the specific conditions the recommendation targeted, such as cell occupancy; and, within Change, Institutionalisation by whether that improvement is protected by a standing inspection cycle and a budget line rather than a single funded project.



STEP-CHECKLIST

- ☐ Are **milestones tied to the recommendation target and supported by appropriate evidence**, with institutional self-report distinguished from practice-based or otherwise claim-fit verification?
- ☐ Are **confidence, variation, contrary evidence and evidence gaps stated explicitly**?
- ☐ Is any **contribution claim kept separate from proof of Change and bounded** to what the evidence supports?
- ☐ Has **implementation learning been routed to the relevant Step** rather than treating incomplete progress as failure?



Closing the Loop, Keeping the Inquiry Open

This Guide began from a simple problem. Justice reform often produces large quantities of information without creating a reliable pathway from evidence to institutional change. Monitoring generates observations. Analysis identifies recurring problems. Recommendations are issued. Progress is reported. Yet the connections among these activities are frequently assumed rather than designed, leaving reform vulnerable to fragmented diagnosis, generic recommendations, activity-based claims of success and repeated cycles of rediscovery.

The value of the engine does not lie in the novelty of any single Step. Many of its constituent methods are already familiar to trial monitors, legal analysts, evaluators, policy designers and institutional reform practitioners. Its contribution lies in bringing those methods into one coherent structure and insisting that none of them should operate alone.

The framework is therefore less a sequence of technical tasks than a discipline of institutional reasoning. It asks the practitioner to make choices explicit: what is being examined, through which standards, based on what evidence, with what limits, toward what change and through which verification pathway. It does not eliminate professional judgement but it makes it visible, contestable and capable of being updated.

Developed through the experience of effectiveness-based trial monitoring, the framework is not confined to one case type, institution or reform field. Its application will necessarily change with the mandate, evidence base, professional standards and institutional setting of the practitioner using it. What should remain constant is the integrity of the chain: a bounded inquiry, a defensible evidentiary spine, an explicit account of inference, a solution architecture linked to practice and a credible method for testing whether change has occurred.

The engine also carries a particular view of reform engagement. Institutions are not passive recipients of findings, and reporting is not the final delivery of an external verdict. Reform develops through structured interaction among evidence, professional judgement, institutional response and public accountability. Constructive engagement does not require the practitioner to dilute findings or negotiate away inconvenient conclusions. It requires assessments precise enough to be acted upon, recommendations grounded enough to be owned and verification designed in a way that can distinguish commitment from change.

The purpose of the policy engine is to make the reasoning behind reform more coherent, the evidence behind it more traceable and each next decision better informed than the last. It will not turn the policy cycle into a straight line — criminal justice reform, like most policy work, has little interest in behaving that neatly. Nor should apparent stagnation be mistaken for stasis. Even when a reform seems stuck, incentives, relationships, leadership, resources, professional expectations, legal interpretations and neighbouring reforms continue to shift, often incrementally but with consequences that are anything but small. In an interdependent system, movement in one part can alter what is possible in another. The engine therefore brings order without rigidity: enough structure to discipline the reasoning, while allowing the practitioner to revise both the picture of the reform landscape and the pathway through it as the system itself changes. The aim is not to tame complexity, but to keep pace with it.



► I. Defining Deep Dives

At certain points in the application of the policy engine, the process may require additional diagnostic inquiries to assess a visible pattern, issue, gap, or trend. To address this, the practitioner may conduct a Deep Dive to examine a specific question in greater depth, typically by collecting a limited amount of new information, reviewing material already collected through a distinct lens, or synthesizing several existing sources on the same question.

The main monitoring tasks produce the initial evidence, ideally showing what happens in hearings, what legal acts contain, how cases move through the procedure, what official statistics report, which patterns recur, and where concerns or good practices begin to emerge. Such evidence should help describe institutional practice in observable terms. A Deep Dive builds on this evidence to make the analysis more precise, the explanation better grounded, the recommendation more targeted, or the tracking question more measurable.

Conducting a Deep Dive does not imply that trial monitoring, consultations, statistical review or other sources are deficient; rather, those sources may reveal a question that was outside the scope of the original collection design, or that can be understood only by examining a narrower issue in greater depth. The Deep Dive question should be neutral, focused and answerable using the available or accessible evidence and resources. Any additional effort required should be proportionate to the value of the expected information.

A Deep Dive, by itself, does not prove influence, incapacity or institutional failure. It identifies where the evidence raises questions the system must be able to answer. The correct result may be that the original pattern is supported, narrower than first thought, more complex than first thought, not supported, or still unresolved. A conclusion that the evidence remains insufficient is a valid research outcome.

At the same time, a Deep Dive does not create a separate reform track. It should be used only when a focused additional inquiry is likely to make a material difference to the evidence or to the next decision.

In addition, a Deep Dive is not:

- Routine collection of a missing document or piece of information;
- A general expansion of trial monitoring or an unrestricted thematic study;
- An institutional audit, criminal investigation or disciplinary inquiry;
- A substitute for weak or incomplete work in the current Step; or
- Research designed to confirm a conclusion already reached.

► II. Deep Dives across the policy engine

A Deep Dive may be triggered from any Step.

Scope – clarify the reform question, boundaries, relevant institutions or feasibility of the evidence strategy;

Collect – address a source gap; test whether an issue is captured by the collection design; or examine material not covered by the original instruments;

Analyse – test, disaggregate or examine an emerging pattern at greater resolution; as noted above, this may emerge as the most common trigger;

Explain – examine competing accounts or obtain evidence needed to strengthen the explanatory account;

Recommend – test assumptions, feasibility, institutional responsibility or the relevance of comparative options; and

Track – investigate stalled implementation, uneven uptake, unintended effects or an apparent but otherwise unexplained change in practice.

The Deep Dive should return to the Step whose work it was intended to strengthen. It does not automatically move to Explain in every case; the same logic applies regardless of the trigger and return points.

► III. Triggering a Deep Dive

Deep Dives are neither required nor always useful. Trigger them only when the existing work reveals a material question that the ordinary work of the current Step cannot adequately address.

A Deep Dive may be warranted where:

- A pattern is visible but not adequately understood;
- Different sources point in different directions, and the discrepancy cannot yet be explained;
- Statistical or administrative data require disaggregation or interpretation;
- An important subgroup, case type or institutional practice requires closer examination;
- A proposed recommendation depends on an assumption that requires further examination; or tracking shows stalled implementation, uneven uptake or an unexpected result.

Before triggering a Deep Dive, ask:

- Is there a material unanswered question?
- Can it be resolved through ordinary work within the current Step?
- Is the question sufficiently focused and answerable?
- Is the additional research likely to affect the analysis, explanation, recommendation or tracking decision?
- Is the effort proportionate to the value of the answer?

If these questions do not support a focused additional inquiry, the practitioner should remain within the existing Step rather than launch a Deep Dive.

► IV. Shaping the Deep Dive question and sources

To conduct an effective inquiry and avoid the trap of an expansive or undefined question, the Deep Dive question must be narrow.

To define the question properly and narrowly, the following elements should inform the development of the Deep Dive task:

- The pattern or evidence gap that triggered the inquiry;
- The specific diagnostic question;
- The sources of evidence, for example, cases, acts, hearings, literature, or datasets to be examined;
- The limits of those sources of evidence, namely, what they can and cannot support;
- The expected contribution to the relevant analytical or policy decision.

In reviewing the wider issues identified, a Deep Dive would refocus on a specific element.

For example, if the question is how effectively serious organized crime and corruption cases are processed, the Deep Dive may, depending on the data collected, ask:

- How are money-laundering charges framed, tested and decided?
- Are confiscation and asset-recovery measures used in a way that is consistent with the financial features of a case?
- Do plea agreements and resulting decisions comply with the applicable legal standards on the factual basis, sanctions and reasoning?
- Are hearing schedules, judgements and decisions publicly available in practice?
- Do appeal decisions give instructions that first-instance courts can put into practice?

When routing back to the Step and having used the Deep Dive to assess whether the existing evidence holds or to fill critical gaps, the practitioner should then state, briefly and transparently:

- The question examined and the additional evidence considered;
- What the Deep Dive confirms, qualifies, contradicts or leaves unresolved;
- The principal limits of the evidence;
- Whether any earlier finding should be revised;
- Which Step the Deep Dive results return to; and
- What decision or next action the results now support.

Once the practitioner completes both the standard evidence collection and the narrower Deep Dive analysis, they review the two streams of evidence together. The Deep Dive does not override or replace trial-monitoring findings, consultation evidence, statistical or administrative data, or legal and documentary analysis. The respective sources must remain clear and visible. From this point onward, the material is consolidated into the evidence base, with the source and limits of each component clearly set out.

Albers, P. [The assessment of court quality: hype or global trend?](#)

An Overview of the Stages (Heuristics) Model as a Public Policy Analysis Framework, [Onur Kulaç, Hüseyin Özgür](#), 11 May 2017, European Scientific Institute, ESI ([328025887.pdf](#)).

Bencze, M. and Ng Gar Y. 2018. [How to Measure the Quality of Judicial Reasoning](#), Springer.

Better Regulation Guidelines, EC, November 2021 ([swd2021_305_en.pdf](#))

Better Regulation toolbox, EC, July 2023 edition ([BR toolbox - Jul 2023 - FINAL.pdf](#)).

Carothers, Thomas, ed. Promoting the Rule of Law Abroad: In Search of Knowledge. Carnegie Endowment for International Peace, 2006, <https://doi.org/10.2307/j.ctt6wpk74>

CCJE. 2008. [Opinion no. 11 on the quality of judicial decisions](#).

CCJE. 2014. [Opinion n° 17 on the evaluation of judges' work, the quality of justice and respect for judicial independence](#).

[Denis Fougere, Nicolas Jacquemet](#), Policy Evaluation Using Causal Inference Methods, In book: Handbook of Research Methods and Applications in Empirical Microeconomics (pp.294-325), Chapter: 11, Edward Elgar Publishing (([PDF](#)) [Policy Evaluation Using Causal Inference Methods](#)).

European Commission, [EU Justice Scoreboard](#).

Ginsburg, T. 2011. [Pitfalls in Measuring the Rule of Law](#), Hague Journal of the Rule of Law, 3, 2.

Jensen, Erik G., and Thomas C. Heller. 2003. [Beyond Common Knowledge: Empirical Approaches to the Rule of Law](#). Palo Alto: Stanford University Press

Joana Sousa Lourenço, Emanuele Ciriolo, Sara Rafael Almeida, and Xavier Troussard; Behavioural insights applied to policy: European Report 2016. EUR 27726 EN ([JRC Publications Repository - Behavioural Insights Applied to Policy - European Report 2016](#)).

Kleinfeld, Rachel. [Competing Definitions of the Rule of Law: Implications for Practitioners](#). Washington, DC: Carnegie Endowment for International Peace, 2005.

Lee Epstein and Gary King, 'The Rules of Inference', 69 University of Chicago Law Review (2002).

Mark A. Hall and Ronald F. Wright, 'Systematic Content Analysis of Judicial Opinions', 96 California Law Review 63–122 (2008).

Mohr R. and Contini F. 2007. [Judicial Evaluation in Context: Principles, Practices and Promise in Nine European Countries](#).

NCSC. 2013. [International Framework for Court Excellence](#).

ODIHR, 2012. [Legal Digest of International Fair Trial Rights](#).

ODIHR. 2012. [Trial Monitoring – a Reference Manual for Practitioners](#).

- OECD (2025) Public Policy Evaluation Toolkit ([Implementation Toolkit for the OECD Recommendation on Public Policy Evaluation \(EN\)](#))
- OHCHR. 2006. [Rule-of-law tools for post-conflict states - monitoring legal systems](#).
- Open Society Institute. 2008. [Promoting Prosecutorial Accountability, Independence and Effectiveness](#).
- OSCE. 2016. [Handbook on Combating Corruption](#)
- RAND. 1976. [Indicators of Justice: Measuring the Performance of Prosecution, Defense, and Court Agencies Involved in Felony Proceedings – A Guide to Practitioners](#).
- Reza Banakar and Max Travers (eds.), [Theory and Method in Socio-Legal Research](#) (Hart Publishing, 2005).
- Rovaniemi Court of Appeal. 2006. [Evaluation of the Quality of Adjudication in Courts of Law - Quality Project of the Courts in the Jurisdiction of the Court of Appeal of Rovaniemi](#).
- Simion, K. 2016. [Qualitative and Quantitative Approaches to Rule of Law Research](#). INPROL.
- Theory of Change Guidelines, Hivos, 2015, ([hivos_toc_guidelines.pdf](#)).
- Vera Institute of Justice. 2003. [Measuring Progress Toward Safety and Justice](#).
- World Bank, [New Directions in Justice Reform](#) (2012).

