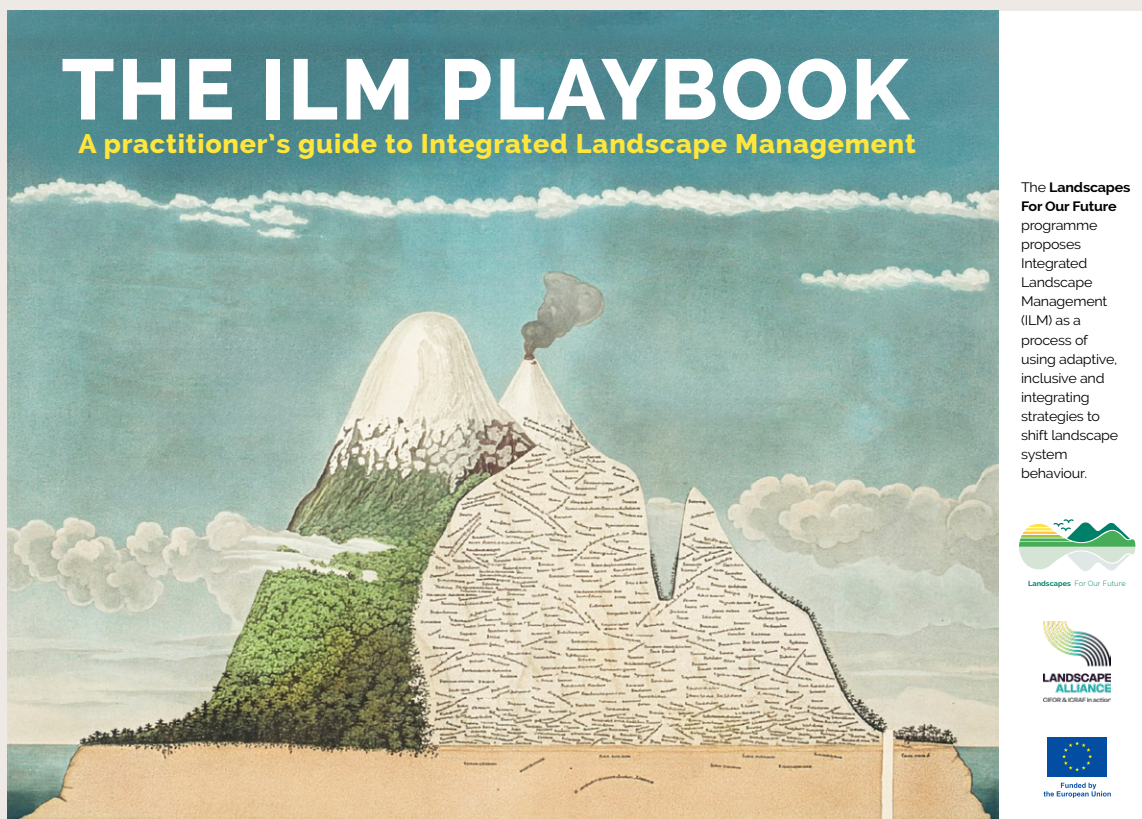


Extract from The ILM Playbook

The ILM Playbook is a practitioner curriculum developed through the EU-funded Landscapes For Our Future programme. Its **interactive web edition** provides a quick, non-linear route through the concepts and tools, while the **full-length PDF** provides the conceptual foundations, evidence and methodological depth behind them.

The concepts, methods and practical guidance were developed through the Central Component's work with project teams, partners and practitioners across the programme, and through the experience and learning generated in participating landscapes.



Developed by

Landscapes For Our Future Central Component, Landscape Alliance

Lead writing and synthesis

Kim Geheb

Conceptual development, review and editorial contributions

Khalil Walji, Mieke Bourne and Dominique le Roux (Landscape Alliance); Marcela GalvisHernandez (FAO Colombia); Camila Alvarenga, Lucimara Marcelino, José Luís Monteiro and Rogério Rosa (Oikos)

Design and production

Dominique le Roux and Elinore de Lisle

With contributions from

Landscapes For Our Future project teams and partners across 36 countries

Explore the full interactive ILM Playbook or access the complete reading edition at:
landscapesfuture.org/ilm-playbook

Monitoring, learning, & evaluation

Key take-aways

- We use MLE – monitoring, learning, *and* evaluation, with learning at the centre – because what conventional M&E reads as failure (“we did not do exactly what we said”) is in complex systems the most valuable form of information.
- Monitoring in a complex system is the systematic observation of how the system responds to your foments – primarily through the three maps, plus attention to ripples, dampening/amplifying loops, cascade signals and absence.
- We distinguish *compliance* indicators (what you did), *movement* indicators (whether the system is shifting towards the alt-heading) and *learning* indicators (whether the team’s read of the system is deepening) – and treat learning as the primary measure of project quality and sincerity.
- The *proposition log* – original proposition, what actually happened, what was learned, how the next foment was redesigned – is the intervention’s learning history and the document a successor team would need to continue the work.
- Three audiences need different things: the team needs real-time learning, the donor needs narrative reporting (more informative than indicator dashboards), and the system needs documentation that outlasts the intervention.

Why MLE, not M&E

Conventional monitoring and evaluation asks: did we do what we said we would do, and did it produce what we predicted? In a complex system, the answer is almost always ‘no, not exactly’ – which conventional M&E interprets as failure rather than as information. The Playbook’s release-observe-adjust cycle is a learning architecture, but conventional M&E cannot see it because it is looking for compliance with a pre-specified plan.

We use MLE – monitoring, learning, and evaluation – with learning at the centre. This is not a cosmetic rebranding. It reflects a substantive reordering of priorities. Monitoring is not indicator-tracking; it is the systematic observation of what the system does in response to foments. Learning is the team’s capacity to revise its read of the system and redesign foments accordingly – the whirlpool described in the Adaptivity section. Evaluation is the honest assessment of whether the intervention is producing movement towards the alt-heading, conducted not for a distant donor but for the team and the system.

MLE serves three audiences, each of which needs different things:

- **The team** needs real-time learning to assess how the system has responded to its foment. The question the team asks is: what did the system do in response to what we released, and what should we try next? This is the primary audience. If MLE does not serve the team, it serves nobody.
- **The donor** needs assurance that money is well spent and the intervention is credible. The question the donor asks is: is this intervention making progress, and is the team competent? We address this in the Guidance for Donors section, but the short answer is that narrative reporting – what was released, what the system did, what was learned, what will be tried next – is more informative than indicator dashboards and more honest about what the intervention is actually doing.
- **The system** needs documentation that outlasts the intervention – what was tried, what the system did, what was learned. The question the system asks (through future practitioners, successor projects, and the communities themselves) is: what do we now know about this landscape that we did not know before? This audience is routinely ignored, but it is the most important for long-term impact.

What monitoring looks like in a complex system

In conventional project management, monitoring means tracking indicators against a plan. In a complex system, monitoring means observing the system's response to your foment. The distinction is fundamental. Indicator-tracking asks whether the intervention is on schedule. System observation asks whether the intervention is producing movement – and if so, in which direction?

The Playbook's primary monitoring instruments are the three maps described in the Sensing and Intervention Design and Implementation sections: the **strand map**, the **power vista**, and the **practice map**. These are living documents – updated after each cycle of release-observe-adjust. If the maps have not changed in six months, either the team is not sensing or the system is not responding. Both are problems. If the maps have changed in ways the team did not predict, that is not a failure – it is the most valuable kind of monitoring data, because it reveals something about the system that was not visible before.

In addition to the three maps, monitoring in a complex system attends to:

- **Ripples:** is the disturbance from your foment travelling? Who is talking about it? What adjacent practices are shifting? Ripple monitoring is inherently qualitative – it involves listening, observing, and asking, not counting.
- **Dampening and amplifying:** is the system absorbing your foment (dampening) or is the effect growing (amplifying)? If dampening, which regulatory forces are doing the absorbing? This tells you where the system's resistance lies – and where your next foment needs to target.
- **Cascade signals:** are actors who were not directly involved in the foment beginning to respond? A cascade is the most powerful sign that the system is shifting, but it is also the hardest to attribute – which is why conventional indicator-based monitoring cannot see it.

- **Absence:** what is not happening? Which actors are not responding? Which interaction spaces remain unchanged? Absence is monitoring data – it tells you where the system’s regulatory architecture is strongest and where your foments have not yet reached.

On indicators

Indicators are reductionist. They reduce the irreducible complexity of a landscape system to a number, and then treat that number as though it tells you something meaningful about change. In our experience, there has rarely been an instance where an indicator actually indicates anything particularly useful about system-level change. This is not because the people designing indicators are incompetent. It is because indicators are structurally ill-suited to complex systems.

The problems are well documented. Indicators decontextualise: a number stripped of its story tells you nothing about why it changed, whether the change is durable, or whether it matters. Indicators incentivize the wrong behaviour: teams design activities that move the indicator rather than the system, because the reporting instrument rewards the former. Indicators are vulnerable to manipulation, masked by a widespread belief that quantitative data are more trustworthy than qualitative data – a belief for which there is no evidence. And indicators create a yawning gulf between what the implementing team sees as satisfactory on-the-ground progress and what a distant donor demands in numerical form.

We do not reject indicators entirely. Some things can be usefully indicated – particularly compliance-level information (how many workshops were held, how many people attended, what proportion of the budget has been spent). But it is important to distinguish between three types:

- **Compliance indicators** measure whether you did what you said you would do. They are useful for fiduciary accountability but tell you nothing about the system. Most conventional indicators are compliance indicators disguised as impact indicators.
- **Movement indicators** measure whether the system is shifting towards the alt-heading. These are harder to define, harder to measure, and vastly more important. They include: changes in the three maps (are strands shifting weight? are interaction spaces reconfiguring? are practices changing their MMC composition?); evidence of ripples and cascades; shifts in the narrative landscape (are different stories being told?); and changes in actors’ causal beliefs about their own system (see below on ChaRL’s psychometric approach).
- **Learning indicators** measure whether the team is learning. This is the most important category and the one conventional M&E cannot see. Learning indicators include: the rate at which the team’s read of the system is deepening (are the maps getting richer and more accurate?); the speed at which failed foments are discarded and redesigned; the extent to which learning from one cycle feeds into the design of the next; the quality of the team’s sensemaking (are interpretations becoming more nuanced?); and the degree to which cross-team learning (the curation function from Team Design, Principle 7) is functioning. Learning is an indication of project quality and sincerity. A team that is learning fast is a team that is doing this work properly, regardless of what the compliance indicators say.

Learning as the primary measure of quality

The Playbook places learning at the centre of its epistemology. The release-observe-adjust cycle is a learning cycle. The whirlpool in the Adaptivity section describes how learning spirals inward towards deeper understanding rather than progressing linearly towards a pre-specified outcome. Sensing is learning. Foment redesign is learning. The three maps are learning artefacts. If the team is not learning, the intervention is not working – regardless of what the logframe says.

Three dimensions of learning matter:

- **The pace of learning:** how quickly does the team update its read of the system after each foment cycle? A team that takes six months to process what it learned in the field is too slow for complexity. A team that debriefs the same day, updates its maps the same week, and redesigns its foment the same month is operating at the pace the system requires.
- **The mechanisms for learning:** what internal structures does the team have for converting experience into insight? The Liberating Structures approach ‘What, So What, Now What?’ (described in the Sensing section) provides one such structure. The mind-mapping and crime board approach provides another – a persistent, visual, evolving record of the team’s collective intelligence. The learning curation protocol from Team Design Principle 7 provides the cross-team architecture. Without deliberate mechanisms, learning remains individual and tacit rather than collective and actionable.
- **The extent to which learning feeds into decisions:** does what the team learns actually change what the team does? If the strand map has been updated but the ferment design has not changed, learning is decorative rather than operational. The proposition log (see below) provides a way to track this: for each foment, the original proposition, what actually happened, what was learned, and how the next foment was redesigned in response.

Evaluation approaches fit for complexity

Several evaluation methodologies have been developed specifically for complex systems interventions. The Playbook draws on three in particular:

Outcome harvesting works backward from observed changes rather than forward from intended outcomes. Instead of asking ‘did we achieve what we planned?’, it asks ‘what changed, and can we plausibly connect it to what we did?’ This is particularly suited to landscape work where the most important outcomes are emergent – not predicted at the proposal stage, not specified in a logframe, but real and consequential. The method involves collecting outcome descriptions from multiple sources (team members, community members, partner organizations), verifying them, and assessing the intervention’s contribution. See Wilson-Grau and Britt (2012).

Most Significant Change (MSC) is a narrative-based monitoring and evaluation technique that captures what conventional

indicators miss. Project participants – at all levels – are regularly asked to describe the most significant change they have observed over a defined period, and to explain why they consider it significant. The stories are then collectively reviewed and the ‘most significant’ selected through structured deliberation. MSC is valuable because it surfaces changes that indicator frameworks cannot detect, and because the deliberation process itself produces learning. See Davies and Dart (2005).

Developmental evaluation embeds evaluation in the intervention rather than imposing it from outside. The evaluator is part of the team, participating in sensemaking and contributing to real-time adjustment rather than delivering retrospective judgement. This is structurally aligned with the Playbook’s approach: the evaluator is a team member who helps the team learn, not an external auditor who measures compliance. Developmental evaluation asks: is this intervention adapting effectively to what it is learning? Is the adaptation improving the intervention’s capacity to produce movement towards the alt-heading? See Patton (2011).

Smajgl and Ward’s **Challenge and Reconstruct Learning** (ChaRL) framework offers a further dimension: psychometric measurement of belief change among decision-makers. By eliciting causal beliefs before and after structured engagements, ChaRL can detect whether foments are actually shifting how actors think about their system – which is a movement indicator that no conventional instrument can capture. See Smajgl and Ward (2013).

What to report and to whom

For the **team**, the primary reporting instrument is the three maps and the proposition log. The proposition log records, for each foment: the original proposition (if we do X, we believe Y will happen, because of Z); what actually happened when the foment was released; what was learned; and how the next foment was redesigned in response. Over time, the proposition log becomes the intervention’s learning history – a record of how the team’s understanding of the system deepened cycle by cycle. This is the document that a successor team would need to continue the work.

For the **donor**, narrative reporting is more informative than indicator dashboards. Each reporting cycle should address:

- What foments were released and in which interaction spaces.
- What the system did in response – including surprises, failures, and dampening.
- What was learned – how the team’s read of the system changed.
- How the maps have been updated – what the strand map, power vista, and practice map now show that they did not show before.
- What will be tried next – how the ferment has been redesigned in response to learning.
- An honest assessment of progress towards the alt-heading – including whether the alt-heading itself has been revised.

This format gives the donor something that an indicator dashboard cannot: a credible account of whether the intervention is learning and adapting, which is the strongest available evidence that money is being well spent.

For the **system**, documentation should be designed for a future reader who was not present. The proposition log, the three maps in their final state, and a narrative account of the intervention's learning trajectory constitute the system's inheritance from the intervention. If the documentation would not help a future team member understand the system, it is not evaluation – it is compliance.

References and further reading

Davies R and Dart J. 2005. *The 'Most Significant Change' (MSC) technique: a guide to its use.* [LINK](#)

Gaspar D. 2000. Evaluating the 'logical framework approach' towards learning-oriented development evaluation. *Public Administration and Development* 20(1): 17–28. [LINK](#)

Patton MQ. 2011. *Developmental evaluation: applying complexity concepts to enhance innovation and use.* New York (NY): Guilford Press.

Smajgl A and Ward J. 2013. A framework to bridge science and policy in complex decision making arenas. *Futures* 52: 52–58. [LINK](#)

Smajgl A and Ward J. 2015. Evaluating participatory research: framework, methods and implementation results. *Journal of Environmental Management* 157: 311–319. [LINK](#)

Wilson-Grau R and Britt H. 2012. *Outcome harvesting.* Cairo: Ford Foundation MENA Office. [LINK](#)