

COMPANION PIECE

Where the CDI Lands

A practitioner's field guide to Cognitive Delivery in mixed-method environments

Why this companion exists

Most delivery leads currently have three or four AI delivery methods landing on their desk at once. SPDD published last week. BMAD popular in pockets. Kiro showing up in IDE conversations. The instinct is to pick one and standardise. The instinct is wrong.

The methods are not alternatives to each other. They sit at different layers of the delivery system, and the comparisons that put them next to each other as competitors miss what is actually happening. When delivery leads try to choose between them on feature checklists, they end up with a longer list of decisions than they started with.

The Cognitive Delivery Index sits above all of these methods. It does not measure whether your team has adopted SPDD or BMAD or anything else. It measures whether the governance principles those methods all aim at are being practised substantively. The measurement question is invariant to the workflow choice, and that is the design.

This companion places the named methods on a four-layer model, shows where the CDI sits in that model, and works through three concrete scenarios that delivery leads can use as templates for their own decisions. The scenarios carry the operational weight of the argument. Every claim made in the structural sections is grounded in something a delivery lead would recognise from their week.

"The strategic move is not picking a workflow. It is deciding the operating model that workflows must conform to."

This piece is a companion to the six-paper Cognitive Delivery Framework series and the CDI Companion Guide. It does not require either to be useful. The argument stands on its own.

HOW TO USE THIS COMPANION

Read sections one to four to understand the layer model and how the methods place onto it. Use section five (the three scenarios) when you are deciding how to position your own team, programme, or organisation. Return to sections six and seven when you need a one-page case for the operating-model-first approach.

What is in this companion

01	The conversation on every delivery lead's desk	04
02	The altitude problem	05
03	Placing the methods	07
04	Where the CDI sits	09
05	Three practitioner scenarios	11
	A A single team adopting SPDD inside an existing governed delivery model	
	B A programme of six teams deciding between SPDD and BMAD	
	C A delivery organisation of forty teams bringing coherence to mixed methods	
06	What this means for delivery leaders	21
07	Closing	22
08	About the framework	23

The conversation on every delivery lead's desk

Delivery leads have stopped asking whether to allow AI in delivery. The question that has replaced it is harder. Multiple methods are landing at once, each with a different shape, and a strategy is expected by next quarter.

This is the conversation I am having most often with delivery leaders right now. They are not asking whether to allow AI in delivery. That argument has been settled, mostly by the teams getting on with it whether anyone formally approved it or not.

The question that has replaced it is harder. They have read about SPDD, which Thoughtworks published last week. They have heard about BMAD from a principal engineer who saw it at a conference. Their senior architects keep mentioning Kiro. Spec-driven development comes up in every leadership meeting. And they have to decide what to do with all of it.

The instinct is to pick one method and standardise. That instinct is wrong, and most delivery leads sense it is wrong but cannot quite articulate why. It is wrong because these methods are not alternatives to each other. They are not five competing answers to the same question. They are answers to different questions, and they sit at different layers of the delivery system.

This piece is the practitioner's map. It places SPDD, BMAD, Kiro and the broader spec-driven category onto a four-layer model. It shows where the Cognitive Delivery Index sits in that model. And it works through three concrete scenarios in enough operational detail that a delivery lead could use them as templates for their own decisions.

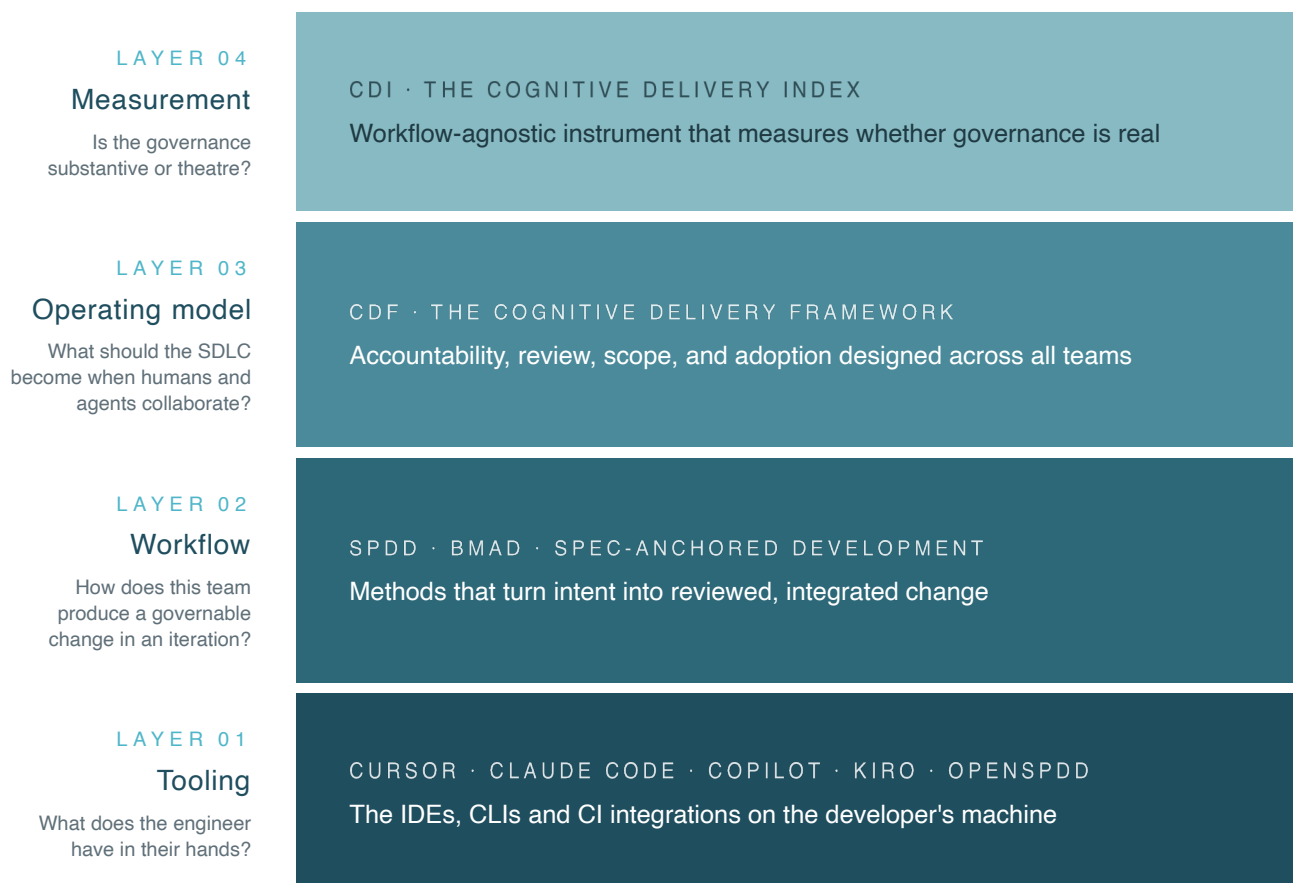
The altitude problem

Method comparisons feel awkward because the methods do not operate at the same altitude. Without a layer model, every comparison reduces to feature checklists, and feature checklists miss what is actually happening.

When delivery leads put SPDD, BMAD, Kiro and the spec-driven movement next to each other, the comparisons feel awkward. Some of the methods overlap in places. Some of them seem to address the same problems differently. Some of them feel complementary but you cannot quite say how. Feature-by-feature comparison reduces the discussion to which one has prompts under version control or which one has named agent roles, and those comparisons miss what is actually happening.

The reason the comparisons feel awkward is that the methods do not operate at the same altitude. A delivery system has at least four layers, and each of them answers a different question.

FIGURE 1 / THE FOUR LAYERS OF AN AI-AWARE DELIVERY SYSTEM



Each named method has a centre of gravity at one of these layers. CDI sits above all of them as a measurement instrument, agnostic to the workflow choice below.

The reason teams confuse the layers is that some methods reach across more than one. SPDD is mostly workflow but ships with tooling (openspdd) and implies governance norms. BMAD is workflow but has views on roles. CDF

is operating model but has views on review mechanics. The reach is real. But each method has a centre of gravity, and once you know what that is, the comparisons become meaningful.

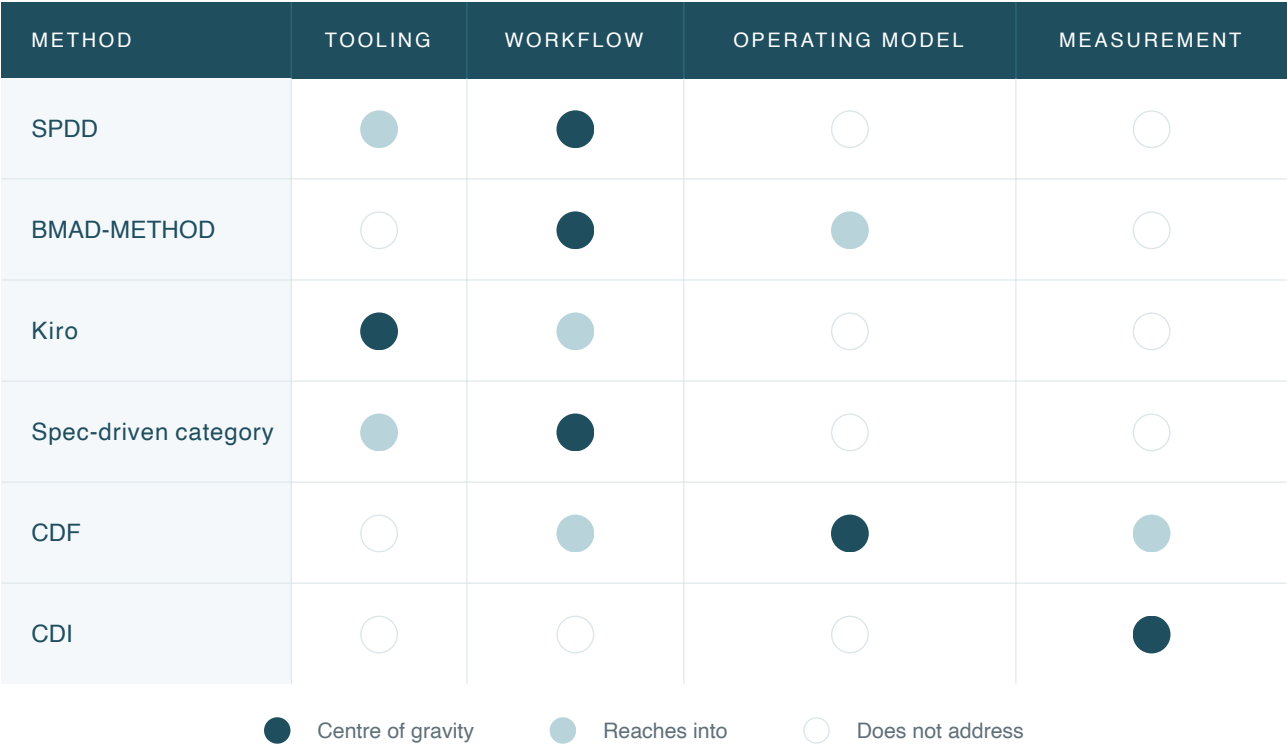
Method comparisons that pretend the layers do not exist always end the same way. They produce a checklist with no architecture behind it, and they leave the delivery lead with a longer list of decisions than they started with. The altitude question is the simplification.

One thing this piece is not. It is not a verdict on which methods are good. SPDD, BMAD and Kiro are doing useful work. The argument is structural, not evaluative. The methods address the workflow question well. They were not designed to address the operating model or measurement questions. That is not a gap in those methods. It is a gap in the discussion that puts them next to each other as if they were competing for the whole stack.

Placing the methods

Each named method has a centre of gravity at one of the four layers. The placement is descriptive, not evaluative. Every method below is doing useful work. The point is to see what each one is for, and what it leaves untouched.

FIGURE 2 / METHOD PLACEMENT ACROSS THE FOUR LAYERS



Structured Prompt-Driven Development (SPDD)

Published in late April by Thoughtworks. Centre of gravity is workflow. Treats the structured prompt as a first-class delivery artefact, version-controlled alongside the code. The REASONS Canvas (Requirements, Entities, Approach, Structure, Operations, Norms, Safeguards) standardises what a structured prompt contains. The closed-loop discipline (when reality diverges, fix the prompt first) keeps prompt and code synchronised. SPDD ships with openspdd, a CLI implementation, which is a deliberate move from method to tooling. What it leaves untouched: organisational accountability, review forum design, adoption staging, measurement. It does not claim to.

BMAD-METHOD

A more agentic workflow, with specialised agents taking on roles like product owner, architect, scrum master, and developer. The discipline is in the orchestration of those agents through a structured agile loop. Centre of gravity is workflow, with a stronger reach into role boundaries than SPDD. What it leaves untouched: how the human roles around those agents are formally defined, how accountability attaches when an agent is the proximate author, and how review of agent-generated change is structured at scale.

Kiro

AWS's spec-first IDE. Centre of gravity is tooling, with a strong opinion about workflow. The IDE is built around steering documents and specs that anchor agent work. It is closer to the engineer's day than SPDD or BMAD because it is the surface they touch. What it leaves untouched: cross-team consistency, programme governance, measurement.

The broader spec-driven category

A growing body of methods, including spec-anchored development as classified by Birgitta Böckeler. The shared instinct is that a clear specification produced before code is the discipline that makes AI-assisted change governable. Centre of gravity is workflow. Each variant differs in what counts as a spec, where it lives, and how tightly it stays in sync with the code.

The thing all of these methods share is the workflow centre of gravity. They are all answering the same fundamental question, which is how a team produces a governable change in an iteration. They differ in their answers, sometimes substantially, but they share the question. What none of them addresses, and what they were not designed to address, is the operating model layer or the measurement layer.

Where the CDI sits

The Cognitive Delivery Index sits above all of the workflow methods. It does not compete with any of them. The measurement question is invariant to the workflow choice, and that is the design.

A team running SPDD can be scored by the CDI. A team running BMAD can be scored by the CDI. A team using Kiro and informal spec-anchored discipline can be scored by the CDI. A team that has invented its own structured prompting practice can be scored by the CDI.

The reason this works is that the CDI does not measure workflow conformance. It measures whether the principles that the workflow methods all aim at are being practised substantively. The six dimensions of the CDI are deliberately orthogonal to method choice.

THE SIX DIMENSIONS OF THE CDI

Provenance Integrity. Are agent contributions traceable in enough detail that accountability can actually attach?

Supervisory Clarity. Is accountability held by a named, identifiable person with defined authority?

Scope Definition. Do agents operate within explicitly bounded domains?

Review Substance. Is review of agent output substantive or procedural?

Constraint Fidelity. Are rules and standards explicit enough to be followed reliably?

Capability Calibration. Are expansion decisions based on evidence rather than enthusiasm?

A team running SPDD with discipline will score well on Constraint Fidelity because the Norms and Safeguards sections of the REASONS Canvas force constraints to be explicit. They might score well on Provenance Integrity because the structured prompts are versioned and traceable. They might still score badly on Review Substance if the team has automated the review of generated code without a human reading it. The workflow choice does not determine the CDI score. The substance of the practice does.

FIGURE 3 / HOW EACH CDI DIMENSION LANDS FOR DIFFERENT WORKFLOW CHOICES

CDI DIMENSION	TEAM RUNNING SPDD	TEAM RUNNING BMAD	TEAM RUNNING MIXED / INFORMAL
Provenance Integrity	Strong by default. Canvas is the artefact.	Variable. Agent logs exist, not always retained.	Weak unless the team has invested deliberately.
Supervisory Clarity	Depends on how the team interprets accountability around the Canvas.	Stronger if PO-as-orchestrator role is defined.	Often the weakest. No method-imposed structure.
Scope Definition	Canvas Norms and Safeguards sections force scope explicit.	Agent role boundaries are defined, scope often is not.	Implicit in conversation, rarely enforced in tooling.
Review Substance	The Canvas helps but does not guarantee real review.	Agent-to-agent review is not human review.	Highly variable. Driven by reviewer behaviour.
Constraint Fidelity	Strong. Canvas demands explicit constraint capture.	Implicit in agent prompt design. Not always retained.	Weak. Constraints often rediscovered each iteration.
Capability Calibration	Method-agnostic. Depends on team and leadership discipline.	Method-agnostic. Depends on team and leadership discipline.	Method-agnostic. Depends on team and leadership discipline.

This is the design intent. A measurement instrument that rewarded a particular workflow would just be a workflow conformance audit. The CDI was designed to be workflow-agnostic precisely so that the question it asks remains the same regardless of which method a team adopts.

"CDF defines the operating model. CDI measures whether the operating model has been instantiated substantively. The workflow methods are how teams do the work the operating model requires."

The practical implication is that an organisation does not need to standardise on one workflow method to have a coherent measurement story. It needs to standardise on what governance has to look like at every layer, and let teams choose the workflow that produces it.

Three practitioner scenarios

The argument so far is structural. The next three scenarios are operational. Each one walks through what changes in the team's week, in the artefacts they produce, in the governance bodies they answer to, in the role boundaries, and in the procurement and tooling reality.

The scenarios are deliberately generic. None of them is a real client. The texture is meant to be recognisable to anyone working inside a governed delivery organisation, public or private sector, regulated or not. Each scenario closes with a CDI snapshot showing what the dimensions look like at the point where the scenario ends.

A NOTE ON THE SNAPSHOTS

The CDI scores in each scenario are illustrative. They show what a credible result looks like for a team in that situation. They are not real assessments. The point of the snapshot is to make the connection between the scenario and the instrument concrete, not to claim a specific outcome for a specific team.

Read in order, the scenarios scale from a single team adopting one method, through a programme deciding between two methods, to a delivery organisation bringing coherence across forty teams running mixed methods. The argument should land more clearly with each step up in scale.

A single team adopting SPDD inside an existing governed delivery model

What changes when a small team picks up a workflow method without disturbing the governance structure around it?

THE TEAM

8 engineers + tech lead, delivery manager, product owner. Internal-facing services.

GOVERNANCE BODIES

ARB · CAB · Security · Audit. Mature, fully governed.

METHODS ON TABLE

SPDD adoption via openspdd. AI assistants ad hoc by some.

DECISION IN FRONT

Adopt or not. Tech lead championing, team sceptical.

The team and the decision

The team. Eight engineers, one tech lead, one delivery manager, one product owner. They build internal-facing services for an organisation that already has a mature governed delivery model: ARB approval for architectural change, change records raised through a CAB for production deployment, security review for any change touching authentication or data egress, audit trace of every code merge.

The decision. They have read the SPDD article and want to try it. The tech lead has installed openspdd on his laptop. The team is sceptical. Some of the engineers have been using AI assistants ad hoc for months and feel they already get the productivity benefit. The tech lead's argument is that the workflow is not for productivity. It is for traceability and review quality.

The mechanics

Week one. They pick a small enhancement to start with. The product owner writes the user story in her usual format. The tech lead runs the SPDD analysis command, generating an analysis context document. Together they review the output. The PO is surprised by how much of the domain reasoning was captured. The tech lead generates the REASONS Canvas. They review it together, refine two of the operations sections, and commit it to the repo alongside the code branch. The engineer assigned to the work generates the code from the canvas. He still writes some of it by hand. The structured prompt becomes the artefact the team reviews together in their planning session.

The artefacts they now produce. Same as before, plus one new artefact: the REASONS Canvas, version-controlled in the repository. The user story is unchanged. The PR is unchanged in form. The ADR is unchanged. The threat model is unchanged. The change record raised at deployment time is unchanged. The Canvas slots in alongside.

What ARB sees. The first time the team brings a change to ARB, the architect on the panel asks how the change was generated. The tech lead is candid. He shows the Canvas. The architect reads the Approach and Structure sections. The conversation that follows is the same conversation ARB always has, but with a clearer starting artefact. The architect remarks afterwards that the Canvas made the review easier than it would have been with just the PR.

Role friction. The engineer who has been using AI assistants ad hoc for months pushes back. He feels the SPDD workflow is heavyweight. The tech lead listens, then explains that the Canvas exists to make the work reviewable, not faster. The engineer accepts this in principle but takes another two iterations to internalise it. By month two, he is generating his own canvases without prompting.

Procurement and tooling reality. openspdd is open source. The team's organisation does not have it on the approved tooling list. The tech lead raises a request. It takes three weeks to be approved, primarily because the tooling team needs to confirm that no team data leaves the local machine. Until approval, the tech lead runs it locally and the rest of the team works in the conventional way. Once approved, it is added to the standard developer image.

CDI snapshot at month three

DIMENSION 01 Provenance Integrity	Level 4	Canvas versioned in the repository. Generation history is present. Some commits still bypass the workflow.
DIMENSION 02 Supervisory Clarity	Level 4	Tech lead is named accountable. Authority clear at team level. Time for supervision is not yet protected.
DIMENSION 03 Scope Definition	Level 4	Agent activity is scoped through the Canvas. Boundaries respected. No technical enforcement yet.
DIMENSION 04 Review Substance	Level 3	Reviews are improving. Inconsistent across reviewers. Some still rubber-stamping.
DIMENSION 05 Constraint Fidelity	Level 4	Norms and Safeguards captured well. Some implicit constraints still missing.
DIMENSION 06 Capability Calibration	Level 3	Team is honest about where the AI struggles. Not yet a structured assessment process.

PROFILE SHAPE — SCENARIO A

01 Provenance Integrity						L4
02 Supervisory Clarity						L4
03 Scope Definition						L4
04 Review Substance						L3
05 Constraint Fidelity						L4
06 Capability Calibration						L3

What this scenario shows

A single team can adopt a workflow method like SPDD inside a fully governed delivery model with very little disturbance to the existing governance structure. The CDI score reflects the discipline of the practice, not the choice of the method. The team is in the upper half of the scale across the board. The Level 4 to 5 gap shows up most clearly in Review Substance and Capability Calibration. That is the team's improvement focus for the next quarter.

A programme of six teams deciding between SPDD and BMAD

What does it look like when a programme realises the standardisation question is smaller than it first appears?

THE PROGRAMME

6 product teams · 6–12 people each. Customer-facing services. One programme director.

WORKING GROUP

Cross-functional. 2 tech leads, principal engineer, delivery manager, head of platform.

METHODS ON TABLE

SPDD vs BMAD. Kiro ruled out as a tooling decision.

DECISION IN FRONT

Recommend a workflow method for all six teams. Strategy due end of quarter.

The programme and the brief

The programme. Six product teams, each between six and twelve people, working under a single programme director. They build customer-facing services. The programme has been told by leadership that they need an AI delivery strategy by the end of the quarter. The programme director assembles a small working group: two tech leads, one principal engineer, one delivery manager, one head of platform.

The brief. The working group's brief is to recommend a workflow method that all six teams will adopt. The candidates on the table are SPDD and BMAD. They have looked at Kiro but ruled it out as a tooling decision rather than a workflow decision.

The mechanics

The first week. The working group does what working groups do. They build a feature comparison spreadsheet. SPDD on one column, BMAD on the other. Twenty-six rows of comparison. After three days, the spreadsheet has not produced an answer. Each method has strengths in different places. The principal engineer observes that the spreadsheet is asking the wrong question, but cannot quite articulate the right one.

The shift in framing. The breakthrough comes when the head of platform asks what the working group is actually trying to standardise. Is it the prompts? Is it the agent orchestration? Is it the review forum? Is it the role definitions? Is it the artefacts? They realise they have been trying to standardise everything at once. Once they separate the question, the answers are different.

FIGURE 4 / WHAT GETS STANDARDISED, WHAT GETS LEFT TO TEAMS

Artefact What the team produces alongside code	Standardised. Every team produces a versioned canvas, REASONS-style.
Review substance What review is expected to do	Standardised. Existing governance bodies, with the canvas as starting artefact.
Agent orchestration How agents are choreographed	Left to teams. BMAD-style if the team wants it. Conventional if not.
Role definitions How human roles shift around agents	Left to teams. Provided accountability is named.
Tooling What runs on the engineer's machine	Left to teams. openspdd, Kiro, BMAD frameworks all permitted.

■ Standardised across the programme
■ Left to individual teams

The artefacts question turns out to have a clear answer. They want every team to produce a structured prompt or canvas as a versioned artefact. SPDD's REASONS Canvas is the cleanest answer to that. BMAD does not quite have an equivalent. The agent orchestration question is harder. Three of the teams have product owners who are excited about the agent role choreography in BMAD. The other three are sceptical. The working group decides not to mandate this layer. The review forum question is left to the existing governance bodies. Both methods are compatible with existing review.

The recommendation. The working group recommends what looks at first like a fudge. Every team will adopt the canvas standard from SPDD. Teams that want to adopt BMAD-style agent orchestration can do so, provided their canvases meet the standard. Teams that prefer to stay with a more conventional workflow can do so, again provided they produce canvases.

The programme director is initially uncomfortable. It does not look like a decision. It looks like a deferral. The principal engineer makes the case that this is precisely the right answer. The programme has standardised on what it actually needs to standardise (the artefact and the review substance) and has left to teams what should be left to teams (the orchestration and the tooling).

Six months later. Three teams have adopted SPDD as written. Two teams have adopted a BMAD-influenced agent setup but still produce canvases. One team has stayed mostly conventional but produces canvases for any change that touches a customer-facing API. The programme produces a quarterly CDI snapshot across all six teams. The teams that scored highest on Review Substance are not necessarily the teams that adopted SPDD most fully. They are the teams whose tech leads modelled substantive review behaviour from the beginning.

Programme CDI snapshot at month six

DIMENSION 01 Provenance Integrity	L3 to L5	Canvas standard has worked. Lowest scores are teams that bypass the canvas for "small" changes.
DIMENSION 02 Supervisory Clarity	L3 to L4	All teams have named tech leads as accountable. Authority is clearer in some teams than others.
DIMENSION 03 Scope Definition	L3 to L5	Highest where the canvas is used most rigorously and where scope is encoded in tooling.
DIMENSION 04 Review Substance	L2 to L5	Largest variation. Driven by reviewer behaviour, not method choice.
DIMENSION 05 Constraint Fidelity	L4 to L5	Strong across the programme. Norms and Safeguards have become standard.
DIMENSION 06 Capability Calibration	L2 to L4	Weakest dimension. No team has structured this yet.

PROFILE SHAPE — SCENARIO B (PROGRAMME RANGE)

01 Provenance Integrity						L3–L5
02 Supervisory Clarity						L3–L4
03 Scope Definition						L3–L5
04 Review Substance						L2–L5
05 Constraint Fidelity						L4–L5
06 Capability Calibration						L2–L4

What this scenario shows

When you have multiple teams, the standardisation question turns out to be smaller than it looks. Most of the choice can be left to teams. The standardisation that matters is the artefact and the substance of review. Method choice falls out of that, not the other way around. The CDI tells the programme director where to focus next: Review Substance variability and Capability Calibration are the programme's two priorities for the next six months. Method consolidation is not.

A delivery organisation of forty teams bringing coherence to mixed methods

What does it look like when leadership asks for coherence across an organisation that is already running every method at once?

THE ORGANISATION

40 teams · ~400 engineers. Single delivery org. Heterogeneous mix.

LEADERSHIP

Director of engineering + 6 EMs, head of architecture, head of platform, CSO.

METHODS IN USE

All of them. SPDD, BMAD-style, ad hoc, homegrown spec-driven.

DECISION IN FRONT

Bring coherence across forty teams. Note: not "standardise."

The organisation and the ask

The organisation. Forty teams, around four hundred engineers, in a single delivery organisation. The mix is heterogeneous. Some teams are well into structured prompting. Some are still using AI assistants ad hoc. Some have adopted BMAD-style agent orchestration. Two teams are running spec-driven development with their own homegrown templates. The organisation has a director of engineering, six engineering managers, a head of architecture, a head of platform, and a chief security officer.

The leadership ask. The director of engineering has been asked by the executive to bring coherence to the AI delivery practice across all forty teams. The phrasing of the ask is suggestive. It does not say "standardise." It says "bring coherence." The director takes that distinction seriously.

The mechanics

The temptation. The temptation, almost immediate, is to mandate a single method. Pick SPDD. Roll it out. Make every team adopt it within a quarter. This is what most delivery organisations do, and it always has the same failure mode. The teams that were doing structured prompting their own way comply on the surface and continue working as before. The teams that were running BMAD-style orchestration are now operating two methods in parallel. The teams that were ad hoc adopt the form without the substance. Six months later, the organisation has compliance. It does not have coherence.

FIGURE 5 / TWO ROUTES THE DIRECTOR COULD TAKE

THE TEMPTATION Mandate one method • Pick SPDD. Roll it out across all forty teams in a quarter. • Treat method conformance as the success measure. • Single audit point. Single training programme. Visible decision. • Failure mode: compliance without coherence. Teams adopt the form, the substance does not change, the surface metrics flatter.	THE CDF APPROACH Define the operating model • Define what governance must look like at each of four layers. • Mandate the substance. Leave method choice to teams within bounds. • Measure quarterly with the CDI. Track shape, not method conformance. • What it produces: coherence in principles, diversity in workflow, intervention based on evidence not enthusiasm.
---	--

The CDF approach. The director takes a different route. He commissions an operating model that defines what governance must look like at each of the four layers, without dictating which workflow method delivers it.

FIGURE 6 / WHAT WAS DECIDED AT EACH LAYER

LAYER 04 Measurement	Quarterly CDI assessments across all forty teams. Aggregated into engineering manager scorecards. Used for support and improvement, not performance management. Not shared with the executive.
LAYER 03 Operating model	CDF principles published internally. Accountability, review substance, scope definition, capability calibration. The papers, lightly adapted, become the internal reference.
LAYER 02 Workflow	Three non-negotiables. Versioned structured artefact for every AI-involved change. Named accountable engineer. Review against explicit norms and safeguards. Form left to teams.
LAYER 01 Tooling	Three approved. openspdd, a BMAD-compatible orchestration framework, and Kiro. Teams can use any of the three or a combination.

The first quarter. The first quarterly CDI snapshot is uneven. Three teams are in the upper half across all dimensions. Twenty teams are in the middle. Seventeen teams are in the lower half. The director does not panic. The first snapshot is a baseline. Improvement targets are set per engineering manager, not per team, to reduce gaming pressure on individual teams.

Twelve months in. The aggregate picture has shifted. The number of teams in the upper half has roughly tripled. The lower half has shrunk. Three teams that consistently scored badly have had targeted intervention. None of these interventions was driven by method conformance. They were driven by the substance of what the CDI revealed.

The methods being used across the organisation have not consolidated. If anything, they have diversified. The director considers this a sign that the operating model is working. Coherence is in the principles and the substance. Diversity is in the workflow.

Organisation CDI snapshot at month twelve

DIMENSION01 Provenance Integrity	28 / 40 at L4+	Strongly improved. Versioned artefact standard now near-universal across the organisation.
DIMENSION02 Supervisory Clarity	32 / 40 at L4+	Strongly improved. Named accountability is the normal practice.
DIMENSION03 Scope Definition	24 / 40 at L4+	Improved. Some teams still operating with implicit boundaries rather than enforced ones.
DIMENSION04 Review Substance	18 / 40 at L4+	Improved but slowest. The Level 4 to 5 gap is still where most teams sit.
DIMENSION05 Constraint Fidelity	30 / 40 at L4+	Strongly improved. Norms and safeguards are now standard practice.
DIMENSION06 Capability Calibration	14 / 40 at L4+	Improvement is slow. This dimension takes the longest to mature, as expected.

PROFILE SHAPE — SCENARIO C (% OF 40 TEAMS AT L4+)

01 Provenance Integrity		70%
02 Supervisory Clarity		80%
03 Scope Definition		60%
04 Review Substance		45%
05 Constraint Fidelity		75%
06 Capability Calibration		35%

What this scenario shows

Coherence is achievable at scale without method consolidation. The CDF as an operating model layer is what makes it work. The CDI is the instrument that tells the director where the organisation is and where to focus. Method diversity is not a problem if the principles are common and the measurement is honest.

What this means for delivery leaders

The strategic move is not picking a workflow method. It is deciding the operating model that workflow methods on your teams must conform to, and the measurement instrument that tells you whether they are conforming substantively. Once those are decided, the workflow choice gets smaller.

This is counter-intuitive for organisations used to standardising delivery practice. The instinct is to mandate the workflow because workflow is what is visible. SPDD is a thing teams can do or not do. BMAD is a thing teams can adopt or not adopt. Operating models and measurement frameworks feel less concrete because they sit above the daily practice. But they are the layers that determine whether the workflow choice produces coherent governance or fragmented practice.

FOUR THINGS TO TAKE AWAY

Workflow methods will keep proliferating. Some will fade. Some will compound. New ones will emerge. A delivery organisation that has hitched its strategy to a particular workflow has hitched its strategy to a moving target. A delivery organisation that has hitched its strategy to an operating model and a measurement instrument has insulated itself from method churn.

Standardise at the artefact and substance level. Every workflow method named in this piece produces some form of structured artefact and some form of review. The differences in what they produce are real but smaller than they look. The differences in whether the practice is substantive are enormous. That is what the CDI measures.

Your delivery organisation already has a CDI score. Whether or not you have measured it. The agent contributions are happening. The review substance is what it is. Measuring it does not change it. Measuring it tells you what to do about it.

Do not aim for Level 5 across the board on day one. The Level 4 to 5 gap is the hardest one to close because it is the gap between the structures existing and the structures being used substantively. Most teams will sit at Level 3 or 4 on most dimensions for a long time. The trend matters more than the absolute level.

Closing

The methods will keep coming. SPDD landed last week. There will be another one next month. The temptation to chase each new method as it appears is one of the things that delivery leaders have to resist. Not because the new methods are bad, but because the strategic question is not at the workflow layer.

The altitude question is durable. So is the measurement question. The Cognitive Delivery Index is the instrument designed for that durability. It does not care which workflow you run. It cares whether the governance is real.

A CONCRETE STARTING POINT

Take one of your teams. Score them against the six CDI dimensions. Be honest about where they actually sit. Pick the two dimensions with the lowest scores. Decide what one improvement would look like for each in the next quarter. Do not change the workflow method. Just change the substance. Score them again three months later.

That is enough to start. The framework is durable. The measurement is honest. The methods come and go.

"Coherence is in the principles and the substance. Diversity is in the workflow."

About the framework

This companion piece sits alongside the six-paper Cognitive Delivery Framework series and the CDI Companion Guide. It is offered freely to delivery leaders trying to make sense of the AI delivery method landscape.

The wider framework

The Cognitive Delivery Framework is published as a six-paper series on Substack. The earlier papers cover the developer progression from Dev 1.0 through Dev 4.0, the redesign of code review for agent participation, the governance principles underpinning hybrid delivery, the question of accountability when agents act, and the adoption tiers organisations move through as agent participation matures. The Cognitive Delivery Index, introduced in paper six, is the measurement instrument.

The book-length treatment is forthcoming, with publication targeted by the end of 2026. Readers interested in early access to material as it is developed can subscribe via the Substack link below.

How to engage with the work

This companion is offered freely. Delivery leaders are encouraged to use it, adapt it to their context, and share what they find. The most useful feedback is the kind that surfaces a layer the model does not capture, or a method placement that fails to discriminate in practice. Both kinds of feedback shape future versions.

If your organisation has used the model to think through a real decision and would be willing to share the outcome, anonymised where appropriate, please get in touch directly. The empirical base for the framework is built one decision at a time.



Jason Cruz Falzon

Creator of the Cognitive Delivery Framework

jason@cognitivedelivery.co.uk

cognitivedelivery.co.uk / cognitivedelivery.substack.com

This companion is a sibling to The Cognitive Delivery Index Companion Guide and the CDI Workshop Toolkit, both version 1.0.