

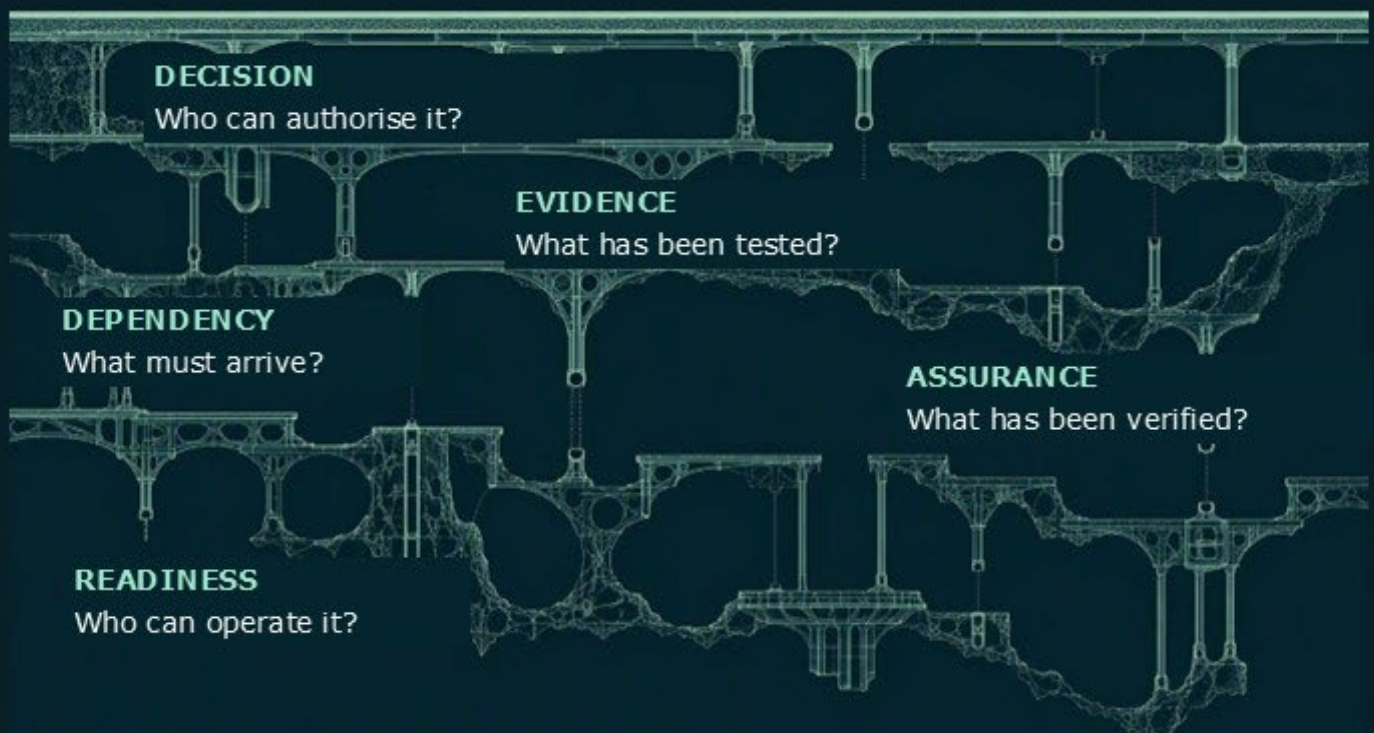
THE DELIVERY CREDIBILITY GAP

An Evidence-Informed Field Guide
to Enterprise IT Delivery

Your project is green.

What can you actually prove?

STATUS: GREEN



Recognise hidden deterioration. Test delivery claims.
Make the next decision useful.

Executive summary

The Delivery Credibility Gap examines a familiar problem: a project can meet its reporting rules while its next delivery commitment becomes harder to defend. A green status may coexist with overdue decisions, fragile dependencies or missing operational evidence.

The thesis is practical and bounded. A useful commitment connects an intended outcome to accountable decisions, coordinated work, proportionate evidence and an operating owner. The PM coordinates those connections within delegated authority. Sponsors, specialists and business owners retain their respective decisions and accountabilities.

The Seven Controls ask whether value remains, useful work moves, dependencies are manageable, decisions arrive in time, capability claims have evidence, operations can run the service and adoption contributes to the intended outcome. They supplement plans, budgets, contracts and required assurance. They are not a complete management standard or seven scientifically independent variables.

Research and professional guidance support attention to early warnings, interfaces, assurance, transition and benefits. They do not validate the guide's named frameworks as predictors of success. The report separates established facts, supported practice, emerging evidence, proposed synthesis and practitioner heuristics.

The practical changes are specific. Show condition, trajectory and confidence separately. Preserve the clock on unresolved decisions. Challenge an avoidable dependency. Ask what a completion claim actually proves and what would invalidate it. Require receiving operators to demonstrate critical tasks. Compare outcomes with a defensible baseline and consider other causes.

Recovery requires a change in delivery conditions and a subsequent test. The 72-hour review is an initial focus period, not a recovery guarantee. AI assistance also needs a defined task, authorised access, accountable review and an evaluation of total effort and quality.

Start with the next consequential commitment. Use one of the nine reusable tools to record the weakest assumption, the person who can act and the evidence needed by the decision date. Retain the method only where it improves the work.

How to use this guide

Start with the control closest to your next decision. Use the toolkit for working records, and the research notes to understand the limits of the claims. All Atlas examples are fictional illustrations, not client case studies.

01 The Project Can Be Green and Still Be Failing	4
02 Manage the System Behind the Plan	5
03 Control 1: Value	6
04 Control 2: Flow	7
05 Control 3: Dependencies	8
06 Control 4: Decisions	10
07 Control 5: Assurance	11
08 Control 6: Readiness	12
09 Control 7: Adoption	14
10 The Signals PMs Notice Too Late	15
11 The Warning Signals I Would Never Ignore	16
12 Stop Reporting Colour. Report Trajectory.	17
13 Recovery Begins With Truth	18
14 The 72-Hour Recovery Review	19
15 Automate Mechanics. Keep Accountability Human.	20
16 How AI Changes the Delivery System	21
17 10 Questions Better Than a Status Report	22
18 The Monthly IT Project Health Review	23
19 The Modern Steering Committee	25
20 The IT PM's Minimum Operating System	26
21 Make the Next Commitment Credible	27

Research method and refinements 28 / Toolkit appendix 29-30 / References 31-33

The Project Can Be Green and Still Be Failing

A green report can describe yesterday's agreement while concealing tomorrow's constraint.

Consider an illustrative cloud migration. The launch date has not changed, spend is inside the approved envelope and every workstream reports green. Yet the identity decision remains open, the connectivity supplier has missed another handover and the recovery rehearsal has moved. The report may be accurate against its definitions. The definitions are no longer sufficient for the decision ahead.

Exploratory early-warning research supports looking beyond formal assessments to how people answer questions and act on concerns [R01]. That does not make intuition a verdict. A hesitant answer is a reason to investigate, not evidence of incompetence or concealment. Ask for the missing artefact, the disputed assumption and the consequence if it remains unresolved.

Separate condition from direction. Condition compares today with an agreed tolerance. Trajectory compares recent observations with a stated horizon. Confidence describes how reliable and complete the evidence is. A project can be within tolerance and deteriorating. It can also be outside tolerance and improving after a credible intervention.

At the next review, select one critical delivery path. Compare the last three observations of its decisions, dependency commitments and verification results. Keep the denominator and scope constant. Record what changed, what did not, and whether the evidence is old. Escalate according to exposure and available authority, not a colour change alone.

Avoid converting every fluctuation into an alarm. A planned discovery phase may legitimately uncover more issues. Investigate whether the team is reducing consequential uncertainty quickly enough for its next commitment. The useful result is an intervention or a justified decision to keep observing.

The Green Project That Is Quietly Dying

REPORTED STATUS		
	GREEN	then and now
Actionable decision age	4 days	→ 18 days
Missed commitments	0	→ 2
Recovery evidence	Planned	→ Deferred

Ask what would have to be true for the green assessment to remain credible.

Basis: R01 | Visual: proposed synthesis; numerical scenarios are illustrative.

Manage the System Behind the Plan

Use the plan to expose constraints, then govern the system that produces the result.

A network rollout can have a detailed schedule and still depend on one security architect approving every site design. Replanning each site does not create more decision capacity. It merely redraws the queue. The PM must connect the schedule to authority, technical interfaces and the evidence needed to move safely.

DORA's work on loosely coupled teams links independent change and testing to delivery capability [R05]. Its context is software. Applying that idea to a data-centre move requires judgment: a shared power cutover cannot be wished away through team autonomy. The useful question is which dependencies are inherent and which were created by organisational design.

The thesis is a proposed synthesis: a useful IT delivery commitment connects an intended outcome to accountable decisions, coordinated work, proportionate evidence and an operating owner. The PM coordinates those connections within delegated authority. Sponsors own investment choices, specialists own their technical judgements and business owners own benefits. Credibility means the claim fits the available evidence; it does not mean certainty or guaranteed success.

Map one outcome backwards. What behaviour must change? What capability enables it? What must be demonstrated before release? Which decisions and interfaces make that demonstration possible? Then map feedback forwards: what evidence would cause a design change, a revised forecast or a stop decision?

Keep integrated milestones, cost control, contracts and required assurance [R03]. Seven Controls is a review lens, not a complete management standard. Funding, capacity, procurement, safety, security and stakeholder obligations cut across it. If a material concern has no owner or record, add it explicitly rather than force it into a reassuring category.

The diagram shows relationships and feedback, not a compulsory sequence. Readiness and adoption work start early. Scale evidence effort to the consequence and reversibility of the next choice; exhaustive proof before every action can itself delay learning. Review the method if it adds reporting without changing decisions or outcomes.

The IT Delivery System

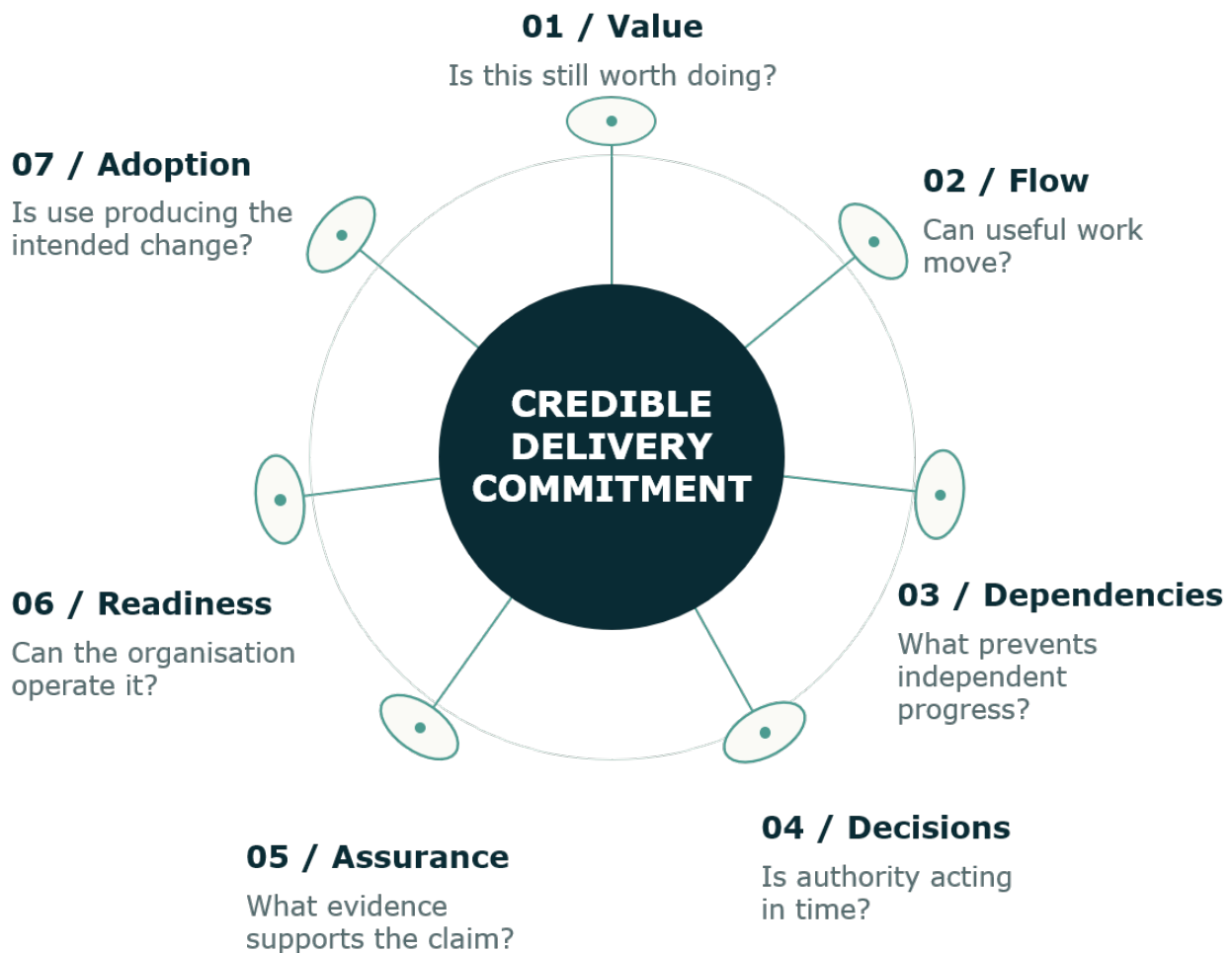


Ask what makes the next commitment credible, and what could invalidate it.

Basis: R03, R05 | Visual: proposed synthesis; numerical scenarios are illustrative.

The Seven Controls of IT Delivery

Seven inspection lenses around a credible delivery commitment.



**Inspect the connections.
Do not calculate an overall score.**

Budgets, capacity, contracts, authority and risk cut across every lens. Keep integrated plans and required specialist assurance.

Basis: proposed synthesis informed by R03, R05, R06, R07 and R08. These are not seven statistically independent predictors or a complete management standard.

Control 1: Value

Is this still worth doing?

An ERP programme discovers that the acquisition it was designed to integrate will not proceed. The baseline remains approved. The investment logic does not. Continuing without revisiting value turns an old authorisation into a substitute for a current decision.

PMI's benefits framework asks organisations to identify, deliver and sustain benefits with accountable owners and measures [R06]. It is practice guidance, not proof that a benefits register causes success. The working method here is to make the chain from problem to measured outcome explicit enough to challenge.

Start with a business problem expressed in observable terms. Identify the capability and the behaviour required to change it. Name a benefit owner who can change the business process and accept the measurement approach. Define a baseline, observation period and material external influences. Avoid translating every benefit into speculative cash savings; resilience, mandatory compliance and service access can warrant explicit non-financial criteria.

At each value review, compare continuing with reducing scope, deferring, stopping or solving the problem differently. Include remaining cost, exit obligations, transition risk and expected outcomes. Already-spent money alone cannot justify another commitment. A cost that is contractually unavoidable still belongs in the option comparison. Monthly is a starting cadence; material changes trigger an earlier review.

An illustrative infrastructure refresh may still be justified when direct savings disappear because an unsupported component creates an unacceptable operating exposure. State that revised rationale openly. Do not preserve an obsolete savings number to avoid revisiting approval.

End the review with a decision, its owner and the next evidence trigger. A benefit estimate without a route to verification is an assumption. An estimate with an owner, a baseline and a test becomes something the programme can manage.

Monthly value review



At each link: what assumption must hold?

Value changed?	Recheck problem and benefit.
Cost changed?	Include exit and transition.
Approve today?	Compare credible alternatives.

Revisit the investment case when its assumptions change.

Basis: R06 | Visual: proposed synthesis; numerical scenarios are illustrative.

Control 2: Flow

Can useful work move through the delivery system?

A migration factory reports a growing number of completed builds while the acceptance queue expands. More build capacity may increase the queue further. The relevant unit is a verified migration that operations can accept, not a server configuration marked done.

DORA recommends small batches that support rapid feedback, and warns against regrouping them downstream [R04]. GAO similarly describes incremental software delivery with continuing evaluation [R17]. Neither justifies forcing every enterprise change into daily production releases.

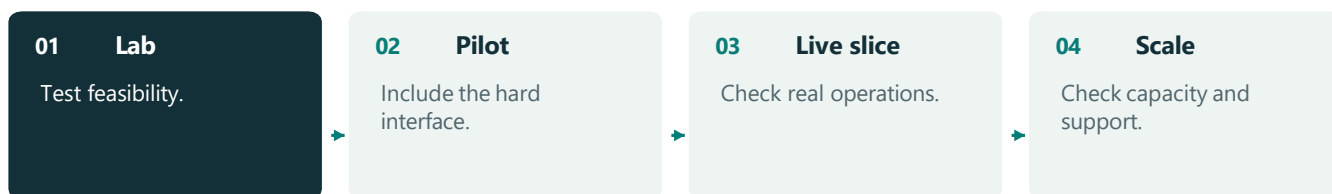
Trace one representative work item from assumption to accepted evidence. Record active work, waiting, handoffs and returns. Use elapsed time alongside effort: a two-hour review that waits twelve days is a different constraint from a twelve-day engineering task. Distinguish deliberate waiting, such as a required observation window, from avoidable queues.

Choose the smallest safe commitment that answers the next material question. A cloud pilot can test identity, monitoring and recovery together for a limited workload. An ERP prototype can validate one end-to-end process before country-specific configuration scales. A hardware programme can validate rack design or factory acceptance while long-lead procurement remains on an integrated schedule.

For each stage, specify the hypothesis, acceptance evidence and what remains untested. Include a difficult but material interface in the pilot, or state its exclusion. Laboratory success does not establish production capacity. Extra decomposition also creates integration and transaction costs; choose the smallest useful test, not the smallest possible task.

Track blocker age, work in progress and rework only where the data changes a decision. Keep work-item boundaries stable enough to compare. If teams split tickets to improve the metric while acceptance remains stalled, the indicator has become a target detached from its purpose. Remove the incentive and inspect the queue.

Commitment and feedback distance



Each expansion needs new evidence. Learning can change the plan.



Shorten the distance between assumption and evidence.

Basis: R04, R17 | Visual: proposed synthesis; numerical scenarios are illustrative.

Control 3: Dependencies

What prevents independent progress?

A cybersecurity rollout cannot test policy enforcement until a shared directory team delivers a group structure. The PM records the dependency and chases an update each week. That may be necessary, but it leaves the delivery design untouched.

Independent testing and deployment are central aims in DORA's discussion of coupling [R05]. Dependency Engineering translates that foundation into a proposed sequence: eliminate, decouple, contract, control. It is not a validated four-stage method, and not every dependency should be removed.

Eliminate asks whether the dependent work is still necessary. Remove a duplicate approval or an unused integration only after confirming its purpose and authority. Decouple asks whether an explicit interface, test double or separate release can permit safe progress. A simulated interface still needs eventual integration validation.

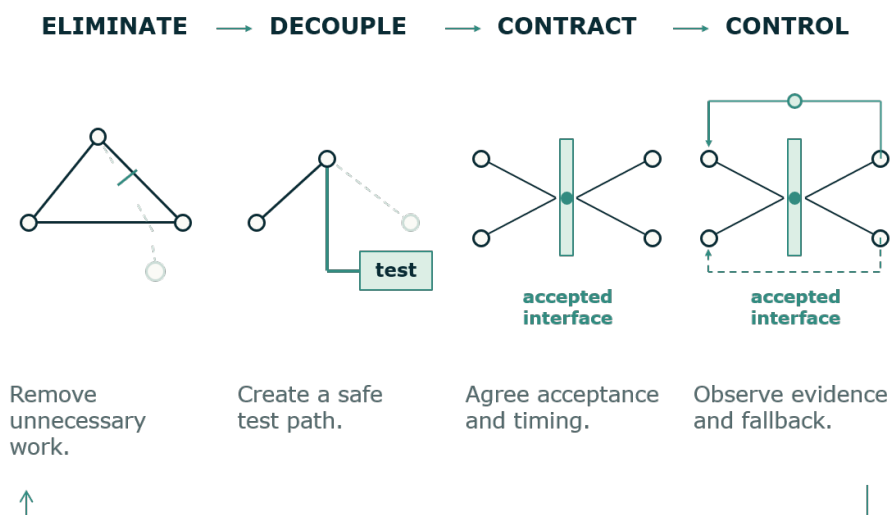
Contract makes the surviving relationship precise: provider, consumer, deliverable, interface

version, acceptance evidence, required-by date and change notification. Here contract means an operational agreement; commercial changes still need authorised procurement and legal review. Control monitors the commitment, its evidence and the fallback when delivery becomes doubtful.

Apply the sequence to a dependency that threatens the next commitment. Compare redesign cost, duplicated work and residual exposure with simply managing it. Shared security controls, utility energisation and integration tests may be necessary coupling. Contract and control apply throughout; the four verbs are options to revisit, not a rule that every dependency must pass through four gates.

Every accepted dependency needs two-sided ownership. A named consumer must verify that the delivered interface works for the consuming service. Delivery by the provider does not automatically mean the dependency is satisfied.

Dependency Engineering



Record the dependency, then challenge the design that makes it necessary.

Basis: R05 | Visual: proposed synthesis; numerical scenarios are illustrative.

Dependencies in five delivery contexts

These are illustrative design choices, not universal prescriptions. Use the canvas to compare the cost of changing the dependency with the exposure of retaining it. Each option needs technical and commercial review in its own context.

APPLICATION / DEPENDENCY ENGINEERING	
Dependencies in five delivery contexts	
These are illustrative design choices, not universal prescriptions. Use the canvas to compare the cost of changing the dependency with the exposure of retaining it. Each option needs technical and commercial review in its own context.	
Cloud migration Primary move: Decouple	What it looks like: Separate preliminary configuration tests from production connectivity acceptance. A lab route may let work continue. What must still be proven: Representative production network, identity and recovery evidence before release.
ERP Primary move: Contract	What it looks like: Define the shared master-data interface before local configuration scales. A test double can expose consumer defects early. What must still be proven: The real data owner can meet quality and timing obligations.
Datacenter delivery Primary move: Control / Sequence	What it looks like: Energisation, cooling and physical access may remain unavoidable coupling. Sequence factory acceptance and documentation work around them. What must still be proven: A genuine site-integration hold point remains before release. Temporary arrangements need owners.
Cybersecurity Primary move: Eliminate + Control	What it looks like: Challenge duplicate approval paths while preserving required segregation of duties. Use pre-agreed policy patterns where suitable. What must still be proven: Unusual exposure still receives the authorised risk decision.
Enterprise application Primary move: Eliminate or Contract	What it looks like: Remove a legacy integration only if the consuming business confirms it is no longer needed. Otherwise version the interface and agree backward compatibility. What must still be proven: The consumer accepts removal or compatibility evidence across the transition.
Provider complete ≠ dependency closed Close the dependency when the consumer accepts the required evidence, or when an authorised redesign genuinely removes the need.	
These examples illustrate context-dependent design choices, not universal prescriptions.	

Control 4: Decisions

Are necessary decisions being made quickly enough?

The architecture team has presented two supported options. Finance owns the cost trade-off, security owns an exception and the programme board assumes somebody else will decide. The item appears as an action for another week while a supplier slot expires.

Decision latency already appears in governance literature and practitioner material; the term is not introduced by this guide [R16]. The proposed operational definition is elapsed time from a decision becoming actionable to an accountable substantive decision. It is a local measure, not a validated universal KPI.

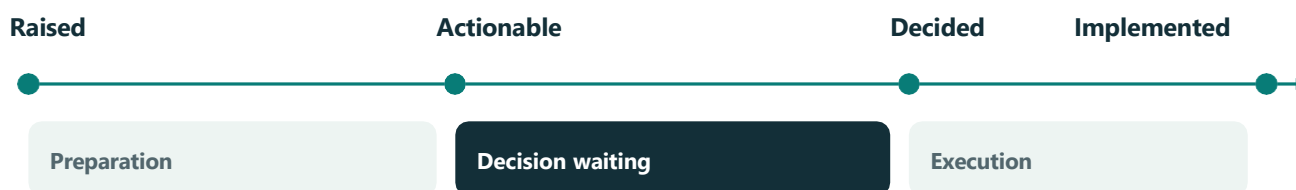
Record first-raised and actionable dates separately. Agree actionable criteria and the authorised decision-maker when opening the item. Preserve the original date if new information arrives; record the changed scope and preparation work. Otherwise a team can improve the measure simply by restarting the clock. The log is an aid to investigation, not a performance target.

For open items, report age as of a named date. For closed items, retain elapsed time and whether the decision arrived before the last responsible date. Use calendar days consistently. Record pauses and reasons transparently; do not silently subtract inconvenient waiting. A deferral needs an owner, next review and exposure statement. It is not substantive resolution.

Prioritise consequence as well as age. A new cutover decision can matter more than an old cosmetic choice. Compare the required-by date with remaining contingency and escalation authority. Separate rejected, superseded and decided items. A deliberate decision to stop can close a choice; postponing the same unresolved choice cannot.

Faster is not always better. Missing safety evidence, a contractual restriction or an irreversible design choice can justify further analysis. Make the cost and benefit of that delay explicit. Finally, check implementation: a signed decision that never reaches the delivery team has not removed the constraint.

Decision-latency timeline



Keep the start date. Record deferral. Check the effect.

Measure waiting without rewarding premature decisions.

Basis: R16 | Visual: proposed synthesis; numerical scenarios are illustrative.

Control 5: Assurance

What evidence supports the claim that this works?

"Disaster-recovery testing completed" can mean the exercise took place, even when recovery exceeded the agreed objective. The report describes an activity. The sponsor needs a bounded capability claim and the conditions under which it was demonstrated.

NIST describes assurance through claims, supporting arguments, evidence and explicit assumptions [R07]. The Delivery Evidence Ladder is this guide's synthesis of that discipline with operational and benefits questions. It is not a NIST model, a maturity score or a hierarchy of research-study quality.

Write the claim before commissioning evidence. Name the service, version, environment, test conditions, acceptance criteria, result and exceptions. Record who reviewed it, when it expires and what change would invalidate it. Ask what observation would contradict the claim. A passed test establishes only what its design and conditions can support.

The six domains answer different questions:

activity happened; an output exists; capability was demonstrated; operations can support use; intended use occurs; an outcome is measured. The final domain also needs a credible account of contribution and costs. A before-and-after improvement alone cannot establish that the project caused the benefit.

Do not let one domain stand in for another. High adoption does not prove security. A technically successful release does not prove business benefit. A benefit can emerge in one increment while another capability remains incomplete. The ladder is a reminder to qualify claims, not a mandatory linear approval ceremony.

For the illustrative migration, replace "DR complete" with "the pilot service was restored within its agreed recovery objectives in the rehearsal environment; production-scale recovery and the out-of-hours operator handover remain unverified." That sentence is less comfortable and more useful. It tells the decision-maker both what is known and what still needs a test.

THE DELIVERY CREDIBILITY GAP

The Delivery Evidence Ladder

01 Activity

The exercise ran

02 Output

A recovery procedure exists

03 Verified capability

Recovery objective demonstrated

04 Readiness

Operators can execute the procedure

05 Adoption

The operating process is used

06 Realised value

Outcome, contribution and costs checked

Six claim domains. Evidence in one does not prove another.

Evidence should be stated at the level actually demonstrated. Activity does not establish outcome, and adoption does not by itself prove assurance or value.

State the claim, its evidence and the conditions where it holds.

Basis: R07, R06 | Visual: proposed synthesis; numerical scenarios are illustrative.

Control 6: Readiness

Can the organisation safely run this tomorrow?

The technical team completes a network change, but the service desk cannot recognise its alerts and the supplier escalation number is only in the project manager's phone. The technology is live. The operating arrangement is incomplete.

Google's SRE engagement guidance describes production-readiness reviews and early engagement [R08]. GovS 002 separately addresses transition and use [R03]. These support involving operations before release; they do not establish one universal readiness checklist.

Use the Day-1 Readiness Test as a practical challenge: if the project team disappeared tomorrow morning, could operations safely run the service? The question is a rehearsal device, not a suggestion to remove support abruptly. Ask the receiving team to demonstrate selected high-consequence tasks using the access, documentation and escalation arrangements they will actually have.

Choose scenarios rather than asking every owner to tick "ready." Can the on-call engineer diagnose a failed dependency? Can the service desk route a user-impacting incident? Can the authorised operator recover data and show its integrity? Can a rollback still be executed after the next irreversible data change?

Every gap needs an owner, due date, evidence requirement and release consequence. Some gaps can be accepted with a time-limited mitigation by the right authority. Others invalidate release. Do not average a critical recovery failure into an otherwise positive score. Separate acceptance of residual risk from a claim that a requirement has passed.

For infrastructure or data-centre work, include facilities, physical access, spares, maintenance windows and supplier availability. For a multi-country application, check local support hours, language and business calendars. Readiness is specific to the service and operating context; copying a checklist without tailoring it can conceal the most important omission.

Two kinds of go-live

Technical release	Operational use
Component installed	Service owner accepts accountability
Test passed	Operators demonstrate response
Monitoring configured	Alerts reach an available responder
Rollback documented	Rollback is feasible and rehearsed

Technical go-live and operational go-live are different claims.

Basis: R08, R03 | Visual: proposed synthesis; numerical scenarios are illustrative.

Make readiness observable

A readiness rehearsal should reveal whether the receiving team can perform the work. Choose a representative incident or recovery scenario and use the runbook, access and escalation route that will exist after project closure.

Assign an observer to record what happened, where assistance was needed and which assumptions failed. Avoid coaching the operator through every step and then calling the operating model proven. Assistance can be appropriate for training, but the claim must reflect it.

State the acceptance decision separately: which gaps block release, which can be accepted within authority, and which require a further test? Keep an expiry date and a monitoring obligation for temporary mitigations.

Readiness area	Demonstration to request
Ownership and support	Receiving owner accepts boundaries; desk routes incident
Observability and access	Alert reaches responder; responder can act
Security and recovery	Exception reviewed; representative restoration checked
Documentation and training	Operator executes current procedure
Supplier and incident response	Escalation works in required support hours
Cutover and rollback	Hold point, recovery path and irreversible steps known

Control 7: Adoption

Is usage producing the intended outcome?

An enterprise application has thousands of logins, but staff still export data to spreadsheets because the approval path is slower than the old process. Login volume is real. It does not answer whether the new capability improved the work.

PMI's benefits framework extends ownership and measurement beyond delivery [R06]. This guide's output-to-value chain makes the intervening assumptions visible: capability must be usable, usage must change relevant behaviour, and that behaviour must contribute to an outcome the organisation values.

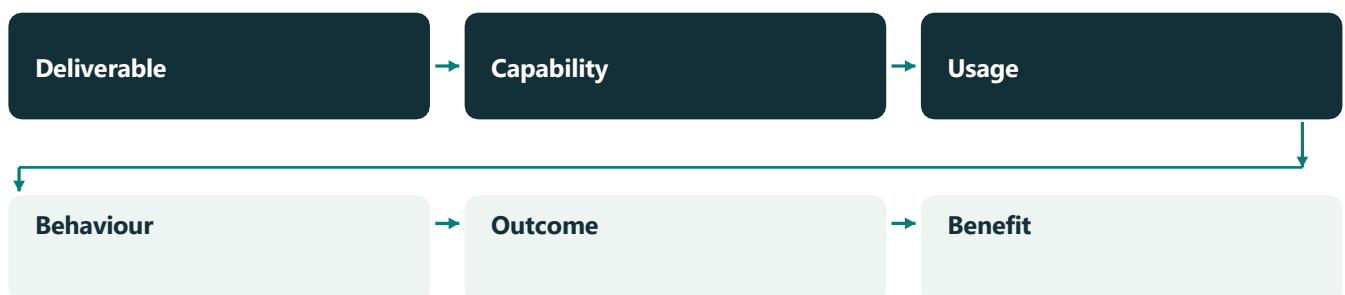
Define adoption for a specific role, task and cohort. For a purchasing workflow, the measure might be eligible orders completed through the intended control path, not accounts created. Record numerator and denominator together. Include exceptions and the reasons people bypass the process. A workaround can be evidence of a design problem, not merely resistance.

Pair adoption with an outcome measure and a balancing measure. Faster processing is less persuasive if error correction or customer effort rises. Compare similar periods and populations where possible, and document other changes that could explain movement. Without a credible comparison, report contribution and uncertainty rather than claiming the project caused the full improvement.

Give the business owner an intervention they can make: revise role permissions, remove duplicate approval, improve training for a specific task, or change the capability itself. A generic communications campaign cannot fix a workflow that prevents users from doing their jobs.

Plan the measurement handover before project closure. Agree who will collect data, investigate drift and decide whether to continue investing. Benefits can outlive the project, but ownership must survive it too. If no one has the authority or capacity to act on the result, the tracker is only an archive.

Output to value



Check contribution, alternatives and costs before claiming benefit.

Go-live does not establish that anything improved.

Basis: R06 | Visual: proposed synthesis; numerical scenarios are illustrative.

The Signals PMs Notice Too Late

Keep risk factors, observed warnings and outcomes separate.

Delivery Vital Signs

CONDITION	TRAJECTORY	EVIDENCE CONFIDENCE
Against a tolerance	Over a stated horizon	In the supporting evidence

Do not average the controls.

Control	Condition	Trajectory	Evidence confidence
Value	Within tolerance	Stable	Medium
Flow	Watch	Deteriorating	High
Dependencies	Intervention required	Deteriorating	High
Decisions	Intervention required	Deteriorating	High
Assurance	Unknown	Unclear	Low
Readiness	Watch	Stable	Medium
Adoption	Watch	Unclear	Low

Low confidence limits the claim. It does not automatically prove failure.

A single integration specialist is a risk factor. Their review queue doubling while commitments are missed is an observable warning. The resulting delayed acceptance is an outcome. Mixing all three in one status field obscures what can still be changed.

Early-warning research describes both formal indicators and softer behavioural signals, with important contextual limits [R01]. It does not provide a calibrated probability of failure for each warning. Delivery Vital Signs therefore uses three separate judgements: condition, trajectory and evidence confidence.

Assess condition against a named tolerance and trajectory over a fixed horizon. Use unknown or unclear when the data cannot support either judgement. For confidence, write a reason: high means relevant, current, corroborated evidence; medium means partial coverage; low means weak, stale or conflicting support. These are proposed local anchors, not calibrated

probabilities.

Record the dated signal, required intervention and its owner. Ask a second reviewer to challenge high-consequence judgements. An unknown may coexist with healthy delivery; its consequence depends on the decision being made.

A green assessment with weak evidence needs qualification, not automatic recolouring. Distinguish what is known about current condition from confidence in the forecast. Record what further observation could change either assessment and the deadline for obtaining it.

Do not aggregate the rows into a weighted number: averaging can hide a release-blocking gap. Discuss disagreements, identify evidence that would change the assessment, and name who will obtain it.

Keep condition, direction and confidence visible as separate judgements.

Basis: R01 | Visual: proposed synthesis; numerical scenarios are illustrative.

The Warning Signals I Would Never Ignore

Translate a warning into an investigation before turning it into a verdict.

A sponsor stops attending. A milestone moves again. Engineers privately doubt the forecast while the official date stays unchanged. Each observation has several possible explanations. What matters is whether someone tests the explanation before exposure grows.

Kappelman and colleagues' IT-project study highlights people and process concerns, including stakeholder support and overloaded expertise [R02]. It is an expert-priority study, not a failure-prediction algorithm. The operational warning list here is a practitioner translation, not a new empirically ranked "top ten."

Start with contradictions: a stable forecast despite evidence loss; completion claims without acceptance; confidence without an available test result. Then inspect accumulating constraints: repeated decision deferral, missed external commitments, growing rework, and critical people allocated beyond realistic capacity. Finally examine displaced safeguards: compressed testing, late operating-team involvement and adoption work removed to protect the date.

Use two questions for each signal. What mechanism could connect this observation to the next delivery commitment? What alternative explanation would make it less concerning? Repeated milestone movement may reflect newly discovered scope rather than poor execution, but it still requires a revised forecast and business decision.

Protect the route for unwelcome information. Ask for specific evidence in a small working session before escalating a personal allegation. Private concern should become a testable issue, not a rumour in an executive pack. Name the organisational condition that needs attention rather than attributing motives.

Prioritise warnings by consequence, persistence and the time left to intervene. A minor anomaly can wait for another observation. A missing cutover rollback capability cannot. Do not manufacture precision by assigning unsupported numerical weights to behavioural signals.

What the dashboard shows / what experienced PMs notice

Dashboard	Observation	Next check
Date unchanged	Validation moved again	Remaining evidence path
Resource plan filled	One SME serves many streams	Actual queue and availability
Testing underway	Exit criteria relaxed	Accepted residual exposure
Actions assigned	Same decision deferred	Authority and consequence

The warning is useful only when it changes what someone checks or does.

Basis: R02 | Visual: proposed synthesis; numerical scenarios are illustrative.

Stop Reporting Colour. Report Trajectory.

Write a status statement that makes the next decision visible.

"Migration: amber" asks the reader to infer the problem, the urgency and the required response. Different executives supply different meanings. The project then spends its steering meeting reconstructing the context instead of deciding what to change.

The reporting pattern proposed here combines condition, trajectory, evidence and intervention. It is informed by early-warning work [R01], but its four-part format is a synthesis. It complements whatever mandatory reporting taxonomy the organisation already uses.

An illustrative replacement is: "Watch, deteriorating. Connectivity acceptance has missed two commitments and the recovery rehearsal is now outside the planned test window. Evidence confidence is high for the missed dates and low for the supplier's new forecast. The sponsor must choose a smaller release or a later launch before the cutover booking becomes non-refundable."

That statement separates observed facts from forecast uncertainty. Attach the dated commitment record and remaining-path assessment. If the claimed schedule impact has not been analysed, say so; do not convert a plausible concern into a precise delay estimate.

Make the intervention executable. Name the decision-maker, options, recommendation, required-by date and consequence of no decision. If no executive intervention is needed, state what the team is doing and the trigger that would change that position. An executive report should not manufacture escalation to justify its existence.

Keep the previous assessment visible long enough to explain movement. Record whether a change reflects new conditions, better evidence or a corrected earlier judgement. All three matter, but they mean different things. The purpose is a shared view of reality that supports action, not an uninterrupted history of reassuring colours.

From colour to a decision

01 Condition

What is true against tolerance?

02 Trajectory

What direction, over what horizon?

03 Evidence

What supports it, with what confidence?

04 Intervention

Who must do what, by when?

A useful report makes the consequence of doing nothing explicit.

Basis: R01 | Visual: proposed synthesis; numerical scenarios are illustrative.

Recovery Begins With Truth

Recovery changes the delivery conditions, not just the approved dates.

A distressed ERP programme adds weekend working and publishes a new plan. The unresolved process design, vendor interface and sponsor trade-off remain unchanged. More effort may produce more rework unless the recovery changes those constraints.

Ward's recovery guidance distinguishes assessment from recovery [R11]. Montealegre and Keil's airport case describes de-escalation that can redirect or end a failing course of action [R12]. These are practice and case-study foundations, not proof of the RESET mnemonic proposed here.

Recognise means acknowledge that normal control is insufficient and secure authority for review. Establish Reality means reconcile scope, actual acceptance, remaining work, commitments and exposure. Surface Causes means test the mechanisms producing repeated failure rather than merely listing symptoms. Execute Interventions means change the relevant scope, sequence, authority, capacity, architecture or commercial arrangement. Test Recovery means observe whether the intervention changes behaviour and evidence.

Keep competing explanations alive. A supplier may be late because its work is weak, because inputs keep changing, or because acceptance is unclear. Each requires a different intervention. Build a short causal account with the people doing and receiving the work, then test it against dated records.

The recovery decision must include reshape, pause and stop. Protect safe ongoing operations and honour legal or contractual obligations while examining alternatives. Do not freeze every activity by reflex; some work reduces immediate exposure and should continue.

Define recovery tests before declaring success. If ownership was the constraint, did decisions close and accepted work resume? If integration was the constraint, did a representative test pass? Look for improvement over an agreed observation window and check that risk was not merely transferred. Keep alternative explanations open when several interventions change together.



Test the change over time. Reopen the cause if recovery stalls.

An approved recovery plan is not recovery.

Basis: R11, R12 | Visual: proposed synthesis; numerical scenarios are illustrative.

The 72-Hour Recovery Review

Use the first three days to frame an intervention and expose what still needs investigation.

A PM inherits a vendor-led programme with contradictory completion figures. The first request is often a new date. Producing one before reconciling the evidence can create another commitment that the delivery system cannot support.

The 72-hour period is a practitioner timebox. No evidence reviewed establishes it as a scientifically optimal review or recovery duration. It applies assessment-first logic [R11] to the first decisions an incoming PM needs to frame. Complex contracts, inaccessible evidence or safety investigations can require longer.

Day 1 establishes reality. Sample the most consequential "complete" items and inspect acceptance evidence with the receiving owner. Build one version of remaining work, decisions, constraints and exposure. Record disagreement and missing information rather than averaging incompatible estimates.

Day 2 examines causes. Follow one failed commitment from request through delivery and acceptance. Identify where information, authority, capacity or an interface broke down. Compare at least one alternative explanation. Separate sunk commitments from avoidable future spend and distinguish a delivery defect from an approval bottleneck.

Day 3 frames the reset decision. Present options, near-term protective actions, evidence needed to choose, accountable owners and the test of recovery. Give a forecast range only where the underlying work and assumptions justify it. Otherwise give a date for completing the forecast investigation.

The output is a one-page decision brief, supported by an evidence index. Its purpose is to make uncertainty manageable and expose the authority required to act. Schedule a follow-up at the point when the first intervention should produce observable evidence. Do not declare recovery simply because the review ended on time.

Three-day review

01 DAY 1: reality

Accepted work, gaps, exposure.

02 DAY 2: causes

Mechanism and rival explanation.

03 DAY 3: reset

Authority, intervention, next test.

At 72 hours, a clear unknown is better than an invented date.

Basis: R11 | Visual: proposed synthesis; numerical scenarios are illustrative.

Automate Mechanics. Keep Accountability Human.

Delegate tasks with a defined evidence and approval boundary.

An AI assistant can turn a long meeting transcript into a convincing action list. If it invents an owner or loses a qualification, the polished output can create a false commitment. The useful control is not simply a person somewhere "in the loop"; it is a named reviewer checking a defined risk before the output is used.

NIST's AI RMF calls for explicit responsibilities and human-AI oversight arrangements [R09]. PMI's June 2026 announcement describes human oversight throughout its AI standard's lifecycle [R13]. The matrix here is a proposed operating rule, not permission conferred by either source.

For meeting synthesis, require links to the source passages and confirmation of actions by the accountable humans. For document comparison, define the authoritative versions and require review of material omissions. For dependency discovery or schedule analysis, ask for candidate findings, assumptions and a reproducible basis; do not let plausible text become an approved forecast.

Scenario modelling can support a choice only if someone checks the input data, calculation and sensitivity. Confidential data must stay within approved processing arrangements. Give the assistant only the access needed for the task, and distinguish drafting a recommendation from changing a system or communicating externally.

Keep risk acceptance, commercial commitments, value decisions, stakeholder conflict, ethical judgement and final trade-offs with authorised people. AI can prepare evidence and alternatives for these decisions. It cannot inherit organisational accountability through a prompt.

Measure the whole workflow: preparation, verification, corrections and downstream use. A fast draft that takes longer to repair is not automatically an improvement. Start with a reversible task, define unacceptable errors and a fallback, then expand only when observed quality and control justify it.

AI delegation matrix

AI role	Examples	Human control
Draft	Minutes, actions, status	Check sources, owners, omissions
Analyze	RAID, schedules, scenarios	Validate data and reasoning
Recommend	Options and trade-offs	Authorized person decides
Never self-authorize	Risk or commercial commitment	Explicit human approval

Delegate a bounded task with named review, permission and a fallback.

Basis: R09, R13 | Visual: proposed synthesis; numerical scenarios are illustrative.

How AI Changes the Delivery System

Test the actual task, its review cost and the quality of the decision it supports.

A team generates requirements summaries more quickly, but the original requirements disagree. Faster summarisation can distribute the disagreement widely in more authoritative language. A second team uses the same capability with controlled versions, source links and a reviewer, and gets a more useful result.

DORA interprets AI outcomes in the context of the surrounding organisation [R10]. That is useful context, not a universal multiplication law. Software-development findings do not establish the effect of AI on steering decisions, schedule forecasts or stakeholder work. Each PM use case needs its own bounded evaluation.

METR reports selection and time-measurement problems in its newer developer-productivity experiment [R14]. This cautions against transporting a single estimate into a new workflow. It does not directly disprove DORA or establish whether AI helps this PM team. Self-reported speed and measured delivery performance are different observations.

Before scaling a PM use case, name its failure path. For status drafting, can unsupported confidence enter the executive pack? For risk discovery, can plausible but irrelevant risks overwhelm material issues? For schedule assistance, can changed assumptions be hidden behind a precise date? Design checks around those errors.

Run a bounded comparison on similar work. Include reviewer time, missed facts, corrections and whether the output improved a real decision. Keep model, prompt, source version and review rules with the result. Where a clean comparison is impossible, describe the observation as provisional rather than attributing every improvement to AI.

Use the result to change the workflow. Sometimes the right intervention is better source data or clearer decision rights. Sometimes the tool adds no useful value. A responsible adoption decision includes the option to stop the experiment.

AI outcomes depend on the workflow

Input condition	Possible effect	Control to test
Ambiguous requirements	Faster inconsistent outputs	Resolve authority and versions
Weak reporting evidence	Polished unsupported claims	Source-linked review
Clear bounded workflow	Potential useful acceleration	Measure quality and total effort

More output is useful only when the delivery system can verify and use it.

Basis: R10, R14 | Visual: proposed synthesis; numerical scenarios are illustrative.

10 Questions Better Than a Status Report

Use a small set of questions to expose the next consequential decision.

A weekly review can cover every workstream and still fail to identify the decision consuming the remaining contingency. The questions below redirect attention to changes in value, constraints, evidence and operating capability. They are practitioner prompts informed by the preceding chapters, not a validated diagnostic instrument.

Choose the questions relevant to the next commitment. Ask for an answer, its evidence and the person able to intervene. "Nothing has changed" should be supported by a recent observation, not copied from last month. An unanswered question belongs in the evidence-gap list with a date and an owner.

Do not use the set as an interrogation script. Give technical and business owners room to explain uncertainties that the questions did not anticipate. Private concerns may need a separate conversation before they can be tested responsibly in a governance forum.

The final question uses 90 days as a thinking horizon, not a forecast standard. Substitute the next irreversible commitment when that is nearer. Ask what assumptions must remain true, which could fail first and what evidence would justify changing direction. A scenario can be useful without assigning an unsupported probability.

Finish by selecting the few actions that can materially alter the trajectory. If the discussion produces ten new reports and no decision, simplify the process. The questions should replace low-value reporting effort wherever permitted, while preserving mandatory records and specialist assurance.

10 questions for the next review

- 01 What changed that could alter the business case?
- 02 Which actionable decision has been waiting longest?
- 03 Which critical dependency has the weakest evidence?
- 04 What important assumption remains untested?
- 05 Where does work arrive faster than it completes?
- 06 Which milestone has moved more than once?
- 07 What are we calling complete without evidence?
- 08 What can operations not yet do for go-live?
- 09 What supports the expected change in user behaviour?
- 10 If nothing changes, where are we heading in 90 days?

Finish the review with the few actions that change the next decision.

Basis: Synthesis from chapters 1-16 | Visual: proposed synthesis; numerical scenarios are illustrative.

The Monthly IT Project Health Review

Make seven separate judgements and agree the response.

The health review is a decision conversation supported by a one-page record. Its purpose is to compare evidence across the Seven Controls without losing the distinct release or investment consequences. The completed example on the next page uses the same illustrative migration as the toolkit.

Before the meeting, ask each control owner for a dated observation, their assessment and the intervention they need. Use shared definitions of tolerance and the review horizon. Distinguish missing evidence from evidence of failure. Invite the receiving operational and business owners, not only the delivery workstreams.

During the review, examine disagreements and dependencies between controls. A decision delay can move a test window, reduce assurance confidence and create a readiness threat. Record that mechanism once and link the affected rows. Do not create three unrelated actions for one unresolved choice.

The monthly cadence is a default heuristic. A stable programme may use its existing cycle; a cutover or recovery period may require more frequent reviews. The cadence must be shorter than the time available to act on material exposure. Do not wait for the monthly meeting when an agreed escalation trigger has already been crossed.

End with named interventions, authority, due dates and expected evidence of effect. At the next review, compare the result with that expectation. If the intervention did not work, reopen the causal explanation rather than merely extending its due date.

Retire indicators that do not influence decisions. The diagnostic is a synthesis for local use, not an additional mandatory reporting layer or a performance ranking of PMs. Review its usefulness with the people who maintain it.

Run the review

01 Prepare Dated evidence and separate judgements	02 Challenge Disagreement, mechanisms, missing proof
03 Decide Intervention, authority and due date	04 Revisit Did the expected effect occur?

The review is complete when the response is owned, not when every cell is coloured.

Basis: Proposed synthesis; R01, R03 | Visual: proposed synthesis; numerical scenarios are illustrative.

Atlas migration health review

Review horizon: the next cutover decision. This sample is deliberately mixed: the service need remains credible while dependency and decision constraints threaten acceptance. The evidence gap in assurance is explicit.

Value: within tolerance, stable, medium confidence. Sponsor reconfirmed the service need; remaining costs need commercial reconciliation. Flow: watch, deteriorating, high confidence. The acceptance queue grew across two reviews.

Dependencies: intervention required, deteriorating, high confidence. Connectivity missed two commitments. Decisions: intervention required, deteriorating, high confidence. Release choice has been actionable for 18 calendar days.

Assurance: unknown condition, unclear trajectory, low confidence. Production-scale recovery has not been demonstrated. Readiness: watch, stable, medium confidence. On-call access awaits rehearsal. Adoption: watch, unclear, low confidence. Pilot usage lacks an eligible-population denominator.

Intervention: sponsor chooses a smaller release or later launch by 25 September. Network owner secures consumer-accepted connectivity evidence. Service owner demonstrates on-call access. Assurance lead commissions representative recovery verification.

Recovery test: at the next review, inspect accepted evidence and whether it restores a credible release path. Do not treat the sponsor decision or a new date as proof that the constraint has disappeared.

The companion T01 record provides blank and completed versions. Keep the underlying evidence links in the working system; the one-page review is an index to judgement and action.

Atlas migration health review

Review horizon: the next cutover decision.

CONTROL	CONDITION	TRAJECTORY	EVIDENCE CONFIDENCE
Value	Within tolerance	Stable	Medium
Flow	Watch	Deteriorating	High
Dependencies	Intervention required	Deteriorating	High
Decisions	Intervention required	Deteriorating	High
Assurance	Unknown	Unclear	Low
Readiness	Watch	Stable	Medium
Adoption	Watch	Unclear	Low

NEXT CONSEQUENTIAL DECISION

Sponsor chooses a smaller release or later launch by 25 September.

Recovery test: At the next review, inspect whether accepted evidence restores a credible release path.

The Modern Steering Committee

Design the meeting around the decisions only this group can make.

A steering committee spends forty minutes hearing workstream updates, then defers the only cross-functional choice because the options were not prepared. Attendance is high. Decision capacity remains low.

Governance standards place decision-making within an explicit framework of roles and authority [R03]. The proposed agenda here is a practical implementation, not evidence that every steering committee should have the same membership or duration. Some regulated programmes need formal assurance and approval records that cannot be replaced by a short discussion.

Use the pre-read for routine progress. In the meeting, start with changed value and trajectory, then address the decisions required. Each request should state the authority needed, viable options, recommendation, consequences, uncertainty and latest useful decision date. If the committee lacks authority, identify the route before the meeting.

Connect critical dependencies, evidence gaps and readiness threats to those decisions. Do not read seven disconnected status summaries. An unresolved supplier acceptance may require a scope choice and an operational mitigation; presenting the mechanism together allows a coherent response.

Reserve time to inspect previous interventions. Did the extra test environment reduce waiting? Did delegated authority close decisions without increasing unacceptable rework? An action marked complete is not enough when the intended effect has not appeared.

Record the rationale, dissent, conditions and accountable implementer. Communicate the result to the people affected and check implementation. A deliberate deferral needs a reason and a next trigger. If the committee repeatedly cannot make, accelerate or improve decisions, review its remit, preparation and delegated authority before adding another meeting.

Modern steering committee

Agenda	Decision output
Value movement	Continue, reshape or stop
Trajectory and dependencies	Change sequence or ownership
Evidence and readiness	Commission test or accept bounded risk
Decisions required	Choose with authority and rationale
Intervention effectiveness	Continue, revise or stop the response

Governance should make decisions possible and their consequences visible.

Basis: R03 | Visual: proposed synthesis; numerical scenarios are illustrative.

The IT PM's Minimum Operating System

Maintain connected records with a clear owner and decision purpose.

A PMO introduces a new diagnostic while retaining every existing log. Teams copy the same dependency into four places and spend the review reconciling versions. The intended visibility creates another coordination problem.

The eight artefacts below are proposed functional views. They can live in one governed system or several existing tools. The test is whether each supports a real decision and has an accountable owner, not whether it exists as a separate spreadsheet. Planning, control and solution-delivery concerns remain complementary [R03].

Connect the records with stable identifiers. A value assumption links to the decision that changes scope. That decision links to the dependency and milestone it affects. Evidence links to the accepted capability and readiness gap. Adoption measures link back to the intended outcome. Preserve source ownership instead of copying long narratives between reports.

Refresh on meaningful events and before the decision the view supports. A decision log needs updating when a decision becomes actionable or is made. A benefits measure follows its observation interval. A milestone forecast changes when evidence changes, even if the reporting cycle has not arrived.

Keep the integrated milestone model. Physical sequencing, long-lead equipment, fixed dates and contractual windows require explicit coordination. The system lens helps test the assumptions behind that model; it does not eliminate scheduling work.

Review the cost of maintaining each view. If an artefact has no identifiable consumer, decision or required record purpose, simplify or retire it through the appropriate governance process. If one critical concern lacks an authoritative record, add only the minimum structure needed to manage it.

Eight artefacts worth maintaining

Artefact / purpose	Decision	Refresh	Bad usage
Value map	Continue / reshape	Assumption changes	Frozen business case
Milestone model	Sequence / commit	Evidence changes	Date without logic
Dependency map	Remove / secure	Commitment changes	Chasing only
Decision log	Choose / escalate	Raised / decided	Deferral = closure
RAID	Treat exposure	New information	List without action
Evidence / readiness	Release / hold	Test or change	Ticks without proof
Adoption / benefits	Improve / reinvest	Measure cycle	Logins = value
Health diagnostic	Intervene	Review / trigger	Average to green

Every artefact should have a decision it exists to support.

Basis: Proposed synthesis; R03 | Visual: proposed synthesis; numerical scenarios are illustrative.

Make the Next Commitment Credible

Create the conditions for a defensible delivery decision.

A delivery commitment is credible when its scope, evidence, uncertainty and accountable decision are visible. That is a working definition proposed by this guide. Evidence can support a decision without eliminating uncertainty; new information can still invalidate a sound earlier judgement.

This guide proposes a delivery-system lens. It does not prove that one operating model fits every IT initiative or that the profession has abandoned its traditional responsibilities. Plans, budgets, milestones and risk controls remain necessary. Their value increases when they connect to decisions, verified capability and outcomes.

The Seven Controls provide a compact way to inspect those connections. Value asks whether the investment remains justified. Flow asks whether useful work moves. Dependencies ask what prevents independent progress. Decisions ask whether authority acts in time. Assurance asks what the evidence supports. Readiness asks whether the organisation can operate the result. Adoption asks whether use produces the intended change.

Choose one practical change next week. Challenge a completion claim with its acceptance evidence. Turn one recurring action into an accountable decision request. Redesign one avoidable dependency. Ask operations to demonstrate a critical task. Revisit one benefit assumption with its business owner.

Observe whether the intervention helped. Record decision timeliness, accepted capability, residual exposure and the effort of maintaining the method. Improvement in a dashboard alone is insufficient. If the method creates delay, hides dissent or displaces an established control, revise it. Keep the parts that improve real decisions and remove the rest.

The delivery-control loop



Turn uncertainty into a decision, then test what the decision achieved.

Basis: Proposed synthesis | Visual: proposed synthesis; numerical scenarios are illustrative.

Research method and refinements

This is an evidence-informed field guide based on a targeted source review completed on 24 September 2026. It is not a systematic review, meta-analysis or empirical validation of the proposed frameworks. Primary research, official standards and first-party practice guidance were prioritised. Search leads were checked against accessible originals; access limits are identified in the references and dossier.

Established evidence refers to well-supported findings or verifiable facts within their stated scope. Supported practice refers to authoritative guidance or context-specific support. Emerging evidence refers to unsettled findings, especially AI research. Proposed synthesis identifies the guide's organising methods. Practitioner heuristic identifies local conventions such as the 72-hour timebox. A standard's existence does not establish causal effectiveness.

The six named frameworks are retained with explicit boundaries. Seven Controls is an editorial coverage map, distinct from PMI's seven performance domains [R15]. Decision latency has prior usage [R16]; this guide specifies a practical clock and log. The Evidence Ladder separates claim domains. Dependency Engineering orders established design ideas. Delivery Vital Signs separates condition, trajectory and confidence. RESET combines assessment and intervention with a test of effect.

The revised diagrams separate six claim domains and show feedback, ownership and cross-cutting constraints. Confidence uses explicit local anchors. Decision dates cannot be silently restarted. Benefit claims need an account of contribution; recovery checks for displaced risk. The separate Concept-Stress-Test records these challenges and the conditions where each method can fail.

No project-failure percentages, universal KPI thresholds or AI productivity effect sizes are asserted. All scenario values are illustrative. No author client history or professional credentials are invented. The supporting dossier records claims, sources, contexts, limits, rejected assertions and final chapter use.

Bounded originality searches found related terminology and foundations. The contribution is the arrangement, practical definitions, examples and reusable tools; no claim of first invention or exclusive ownership is made.

Toolkit appendix

PM-Toolkit.pdf contains an interactive blank form and a completed illustrative example for every tool. PM-Toolkit.html and PM-Toolkit.md provide editable copies. All examples use the same fictional Atlas migration so that decisions, dependencies and evidence can be followed across records.

Use the forms inside existing governance. Replace duplicate records where authorised. Each working record needs a named owner, evidence link, date and decision consumer. The examples show the expected specificity; their dates and thresholds are not benchmarks.

Tool	Use when you need to...
T01 IT Project Health Diagnostic	Choose interventions across the Seven Controls without hiding a critical gap in a composite score.
T02 Decision Latency Log	Expose the waiting time and consequences of an actionable decision.
T03 Dependency Engineering Canvas	Redesign one consequential dependency before accepting repeated chasing as the solution.
T04 Delivery Evidence Ladder Worksheet	State exactly what a completion claim proves and what it does not.
T05 Day-1 Readiness Checklist	Decide whether the receiving organisation can operate the service.
T06 72-Hour Recovery Review	Produce a one-page intervention decision with explicit unknowns.
T07 Modern Steering Committee One-Pager	Prepare and record the decisions that require steering authority.
T08 AI Delegation Matrix	Define a bounded AI-assisted workflow and its human accountability.
T09 Benefits and Adoption Tracker	Test whether use of the capability contributes to the intended outcome.

Put the toolkit to work

Start with T01, the health diagnostic. Choose the most consequential weak control and open only the supporting tool it needs. A decision blockage points to T02. An external interface points to T03. An unsupported completion claim points to T04. A release question points to T05.

Use T06 when the current delivery approach needs a recovery assessment. Use T07 to bring the resulting choice to the authority that can act. Use T08 before delegating a project workflow to AI. Use T09 to connect adoption and outcome evidence to the business owner's intervention.

For an illustrative dependency chain, record DEP-07 in the dependency canvas, link the release choice D-014 in the decision log, and connect both to the evidence required for readiness. The health diagnostic then summarises the same mechanism instead of creating three disconnected narratives.

Agree completion rules for the records. A dependency closes on consumer acceptance or authorised removal. A decision closes on a substantive accountable choice, with implementation tracked separately. A readiness item passes only on the defined evidence. A benefit claim requires a stated observation and its attribution limits.

Keep the cost of the system visible. If a field is never read or never changes a decision, challenge whether it is needed. Retain mandatory and auditable records, but remove ornamental reporting. Review the tools after a few cycles and tailor them to the consequence of the work.

The publication and assets are prepared for local review. The author can adapt the wording and diagrams before any external distribution. The editable source preserves the distinctions between research, practice, synthesis and heuristic throughout.

References

R01 Williams, T., Klakegg, O.J., Walker, D.H.T., Andersen, B. and Magnussen, O.M. (2012). Identifying and Acting on Early Warning Signs in Complex Projects. *Project Management Journal* 43(2), 37-53. DOI 10.1002/pmj.21259.

Relevant: pp. 40-41, 49-51.

www.ntnu.no / [Read source](#)

R02 Kappelman, L.A., McKeeman, R. and Zhang, L. (2006). Early Warning Signs of IT Project Failure: The Dominant Dozen. *Information Systems Management* 23(4), 31-36. Author-hosted paper; expert-priority evidence.

leonkappelman.com / [Read source](#)

R03 Government Project Delivery (2025). Government Functional Standard GovS 002: Project delivery, version 2.1.

Relevant: governance, planning and control, clauses 8.6-8.9.

projectdelivery.gov.uk / [Read source](#)

R04 DORA / Google Cloud (2025). Working in small batches. Updated 8 December 2025. Relevant: methods and common pitfalls.

dora.dev / [Read source](#)

R05 DORA / Google Cloud (n.d.). Loosely coupled teams. Relevant: independent test/deploy outcomes and architecture boundaries.

dora.dev / [Read source](#)

R06 Project Management Institute (2016). Benefits Realization Management Framework. Relevant: Identify, Execute and Sustain, PDF pp. 3-5.

www.pmi.org / [Read source](#)

Source markers identify foundations. Scenario examples, tool fields and proposed diagrams are this guide's synthesis unless stated otherwise.

References

R07 Ross, R., Winstead, M. and McEvilley, M. (2022). Engineering Trustworthy Secure Systems. NIST SP 800-160 Vol. 1 Rev. 1. Relevant: printed pp. 35-36 and 107. DOI 10.6028/NIST.SP.800-160v1r1.

csrc.nist.gov / [Read source](#)

R08 Google (2018). SRE Engagement Model. The Site Reliability Workbook. Relevant: early engagement and production readiness. Practitioner account.

sre.google / [Read source](#)

R09 NIST (2023). Artificial Intelligence Risk Management Framework 1.0. NIST AI 100-1. Relevant: GOVERN 2 and 3.2. DOI 10.6028/NIST.AI.100-1.

airc.nist.gov / [Read source](#)

R10 DORA / Google Cloud (2025). State of AI-assisted Software Development. Official report overview; interpretation used qualitatively. No effect sizes reproduced.

dora.dev / [Read source](#)

R11 Ward, H. LeRoy (2007). Five critical first steps in recovering troubled projects. PMI Global Congress 2007: Asia Pacific. Experience-based conference paper; introductory failure statistics excluded.

www.pmi.org / [Read source](#)

R12 Montealegre, R. and Keil, M. (2000). De-escalating Information Technology Projects: Lessons from the Denver International Airport. MIS Quarterly 24(3), 417-447. Repository abstract consulted; single case.

aisel.aisnet.org / [Read source](#)

Source markers identify foundations. Scenario examples, tool fields and proposed diagrams are this guide's synthesis unless stated otherwise.

References

R13 Project Management Institute (2026). PMI Publishes World's First Global Standard for AI in Project Work. 9 June. Publisher announcement consulted, not full licensed standard.

www.pmi.org / [Read source](#)

R14 Becker, J., Rush, N., Cunningham, T., Rein, D. and Mahamud, K. (2026). We are Changing our Developer Productivity Experiment Design. METR, 24 February.

metr.org / [Read source](#)

R15 Project Management Institute (2025). PMBOK Guide: Eighth Edition. Official product description; November publication date and April 2026 errata noted. Full licensed text not reviewed.

www.pmi.org / [Read source](#)

R16 Crowder, R. (2026). Design Trigger-Based Decision Rights to Reduce IBP Decision Latency. Gartner, 23 June. Public abstract used only to establish prior terminology.

www.gartner.com / [Read source](#)

R17 US Government Accountability Office (2023). GAO Agile Assessment Guide: Best Practices for Adoption and Implementation. GAO-24-105506, 28 November; reissued 15 December. Official overview consulted.

www.gao.gov / [Read source](#)

Source markers identify foundations. Scenario examples, tool fields and proposed diagrams are this guide's synthesis unless stated otherwise.