Grade Program: RAID Logs, GDPR Rails, and Three-Audience Reporting

A multi-year AI research programme with a major operator in Oman survived long enough to ship because of the governance wrapped around the models, not only the models themselves. This whitepaper reconstructs that governance operating model from the engagement archive and reads it as the real risk-management instrument of the programme. The spine is a monthly cadence: daily stand-ups, weekly meetings, and monthly signed reports, all filing into a single RAID log that sorted every item into Risks, Actions, Issues and Decisions and covered both personal and wells data under one discipline. Sitting a full year ahead of that cadence was an Article-28-style data-processing agreement, signed 16 July 2020 in Assen under Dutch law, that fixed the EU-servers-only rails before any confidential log reached a model, a point an earlier chapter of this programme owns in full and we treat here as a pointer. The reporting itself was layered for three audiences at once: domain experts in geology and geophysics reading model-versus-ground-truth evidence, CIO and CDO leadership reading a before-and-after-MLOps maturity picture across eight dimensions, and the CEO, managing director and board reading a Three Horizons growth chart. And the programme was honest about its binding constraint: slow data collection was named the top risk, mitigated by a hard Phase-3 floor of 10 to 15 usable wells, a floor the delivered data cleared only after two of ten received wells were excluded for abnormal static value ranges, leaving eight usable. The claim throughout is that a research programme stays legible, auditable and fundable because of the cadence, the rails, the audience-layered reporting and the named risk, not because the models were good on any single week.

Tarry Singh
Founder & CEO

Narendra Patwardhan
Research Collaborator

AI-GOVERNANCE / GRC / RAID-LOGS / GDPR / DATA-PROCESSING-AGREEMENT / BOARD-REPORTING

CONTENTS

Research is the part of a business that is allowed to fail, which is exactly why it is the part a board watches most nervously. A production line that misses its numbers is a known problem with a known set of levers. A multi-year research programme that misses its numbers on a given month might be dying, or it might be three weeks from a result, and from the outside those two states look identical. The question a board is really asking, meeting after meeting, is not "is the model good yet." It is "can I still see what is happening in here, and can I stop it cleanly if I need to." A research programme that cannot answer that question loses its funding long before it loses its science.

We ran the technical side of a multi-year applied-AI research programme with a major operator in Oman: a set-prediction fracture and bedding detector for carbonate image logs, a vug-quantification pipeline, and well-to-well correlation across a fractured carbonate field. The models and the field metrics are documented at length elsewhere in this record. This paper is about the other half of why the programme stayed alive: the governance operating model that ran alongside the science and kept it legible to the people paying for it. That governance was not an afterthought bolted on for a steering committee. It was a monthly cadence writing into a RAID log, a data-processing agreement signed a year before the first model, reporting layered for three different audiences, and a risk register that named the real constraint out loud. Our argument is blunt. On a research bet, that scaffolding is not overhead around the product. For long stretches of a multi-year programme, it is the product the board is buying.

The cadence is the heartbeat, and the RAID log is what it writes down

The operating model at the centre of this programme was a rhythm, and it was explicit in the December 2022 steering deck rather than implied. Three review clocks turned at three speeds. Daily stand-ups kept the technical team synchronised at the grain of a day. Weekly meetings pulled the week's work into a shape someone outside the team could follow. And once a month a signed report went up the chain, the artefact the board actually read. Nested this way, the fastest clock feeds the next, which feeds the slowest: a month of stand-ups and four weekly meetings compress into one report that a director can absorb in the time they have for it.

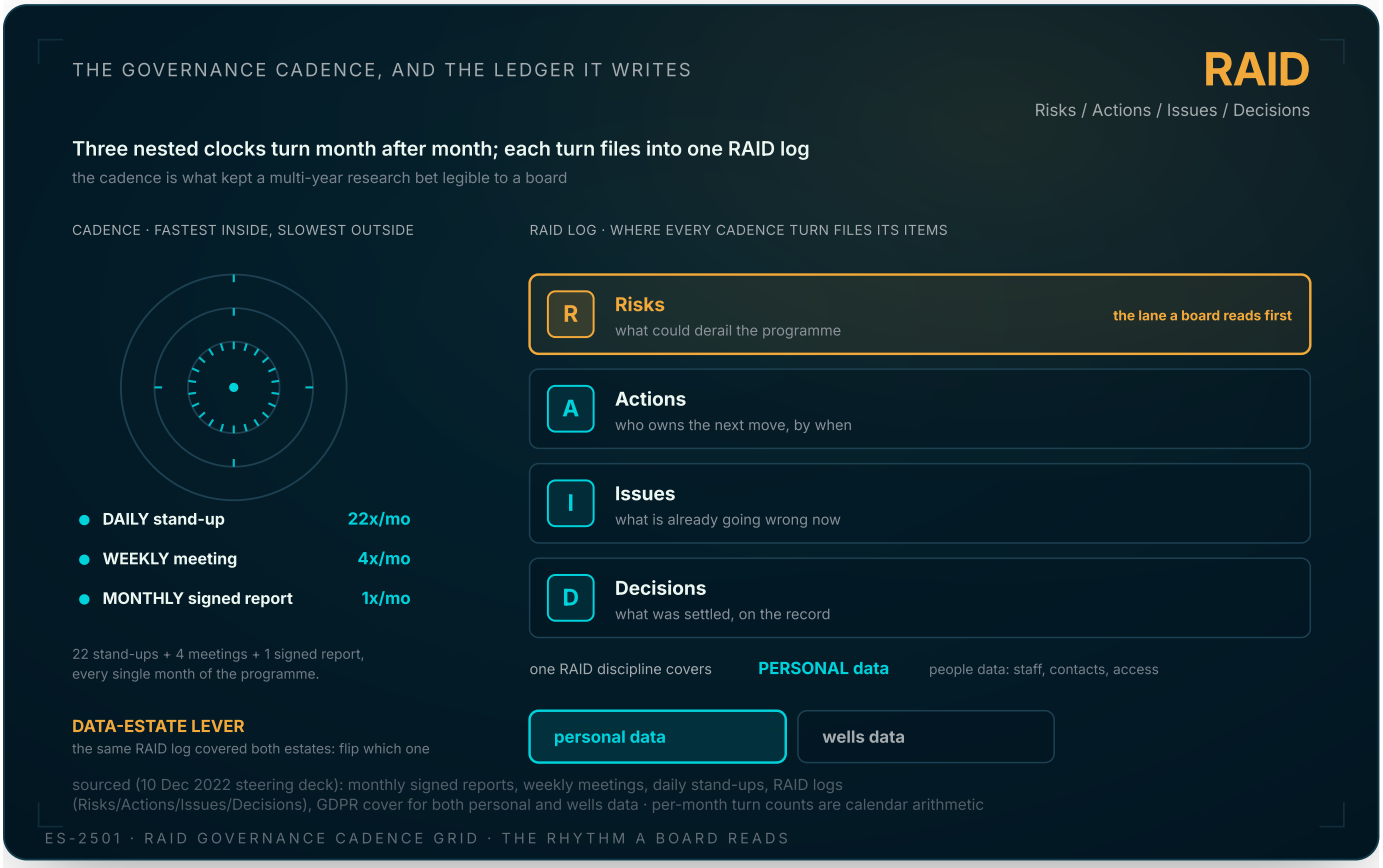
The point of naming the frequencies is that they are not decoration. A programme that only reports monthly has a thirty-day blind spot in which anything can happen and nobody outside the team will know until the report lands. A programme that only stands up daily has perfect local visibility and no altitude, no artefact a board can hold. The three-tier cadence exists so that the same programme is legible at three zoom levels at once, and so that a problem visible in a stand-up on the third of the month is already inside the risk register by the time the monthly report is drafted.

The monthly report was signed, and the signature is not a formality either. A signed report is a person putting their name to a specific account of the month, which changes what gets written into it. An unsigned status update can drift toward the vague and the hopeful because nobody owns the words. A signed one is a claim someone will be asked about at the next meeting. On this programme the monthly report carried the signatures of the technical leads, and that ownership is part of why the cadence produced a governable record rather than a stream of optimistic prose. The board was not reading a newsletter. It was reading a countersigned account it could hold the signers to.

There is an operational cost to running three clocks, and it is worth being honest about it, because the argument that governance is the product falls apart if the governance eats the science. The stand-ups are cheap; they are minutes. The weekly meetings are the expensive tier, because they pull the whole team out of the work to reshape a week into something legible. The monthly report is expensive in a different way, a day or two of synthesis rather than of meeting. What makes the cost worth paying is that each tier feeds the next, so almost nothing is written twice. The stand-up notes become the raw material for the weekly summary, and the weekly summaries become the spine of the monthly report. Run well, the cadence is not three separate reporting jobs. It is one continuous act of keeping the programme legible, sampled at three rates.

What the cadence writes into is the RAID log, and the four letters are the whole discipline. Every item surfaced by any clock is sorted into one of four buckets. **Risks** are what could derail the programme but has not yet. **Actions** are the next concrete moves, each with an owner and a date. **Issues** are what is already going wrong right now. **Decisions** are what was settled, recorded so that a choice made in March is not silently relitigated in July. A RAID log is unglamorous, and that is its virtue: it forces a running programme to keep saying, in four fixed columns, what it is worried about, what it is doing about it, what has

already broken, and what it has committed to. On this engagement the same RAID discipline covered both the personal data of the people involved and the wells data itself, a single ledger over two very different estates.



The governance operating model of a multi-year applied-AI research programme, drawn as a rhythm. On the left, three nested review clocks turn at different speeds, fastest inside to slowest outside: about twenty-two daily stand-ups, about four weekly meetings, and one monthly signed report, every month of the engagement. On the right, the RAID log those cadence turns file into, four lanes for Risks, Actions, Issues and Decisions. The orange element is the only one that argues: the Risks lane, the one that carries the programme's top named risk and the one a board reads first. The data-estate lever flips the covered estate between personal data and wells data, because the sourced discipline covered both. The cadence tiers, the RAID four-bucket taxonomy and the both-estates coverage are sourced from the 10 December 2022 steering deck; the per-month turn counts are the arithmetic of a calendar month, and the clock and lane geometry are presentation, not metrics.

The value of that ledger to a board is not that it catches every problem. It is that it makes the programme auditable in a way that a stream of good news never is. When a director asks, six months in, "what were you worried about in Q1 and what did you do," a RAID log answers in the log's own words, dated, owned, and closed or open. That is the difference between a research programme a board can govern and one it can only trust or defund. Trust is not a control. A RAID log is.

The four buckets are also a discipline against the two most common failures of a research team's own self-report. The first is the risk that never becomes an issue in writing: the team knew for months that data was arriving slowly, but it lived as a shared anxiety rather than a logged Risk with an owner, so when it finally bit, it looked to the board like a surprise. A RAID log forbids that by forcing the anxiety into the Risks column before it becomes a crisis. The second failure is the decision nobody remembers making. Research programmes accumulate choices, which normalisation, which backbone, which wells to exclude, and without a Decisions column those choices get silently reopened whenever a new person joins or a result disappoints. Writing the decision down, with the date and the reasoning, is what lets a programme move forward instead of relitigating its own past every quarter. Neither of these is glamorous. Both are the difference between a programme that compounds and one that churns.

The rails were signed a year before the first model, and that ordering is the point

A governance cadence tells a board what is happening now. It says nothing about whether the programme was ever lawful to run in the first place. For a European vendor handling a Gulf operator's subsurface and personal data, that second question is not a formality, and on this programme it was answered early and on paper. An Article-28-style data-processing agreement was signed on 16 July 2020 in Assen under Dutch law, naming the controller and the processor, fixing that the data transmitted for the work stays on EU private and secure AI servers, and scoping that the data is fed to models and not modified. It named a data protection officer and carried the standard sub-processor-consent, audit, deletion, return and joint-liability clauses.

The load-bearing fact is the date. That agreement was signed roughly a year before the machine-learning engagement's data-and-model phase began. The rails were laid before the first byte of a confidential log moved to a model, not retrofitted once the science was working. That ordering matters because a data-processing agreement signed after the transfer cannot un-send the data; the compliance value of the rails is entirely a function of their coming first. We have written that timeline out in full in an earlier chapter of this programme, and rather than re-derive it here we point to it: the piece on signing a GDPR data-processing agreement before the first model owns the DPA-versus-model ordering, the clause set, and the counterfactual of retrofitting the rails too late. What matters for this

paper is only how that early agreement fits the larger governance picture. It is the foundation the cadence stands on. A monthly RAID log covering wells and personal data is worth little if the underlying right to process that data was never established; the DPA is what makes the whole governance stack defensible rather than performative.

There is a general shape here that outlives this one contract. Compliance scaffolding, on any regulated applied-AI programme, has to lead the data and the model, not trail them. The instinct in a research team is to treat the paperwork as something to finish once the science looks promising, because until then there may be nothing to protect. That instinct is exactly backwards for anything touching regulated data across a border. The rails are cheapest to lay when there is no traffic on them yet, and they are worthless the moment the first unlawful transfer has already happened.

One programme, three audiences, three exhibits

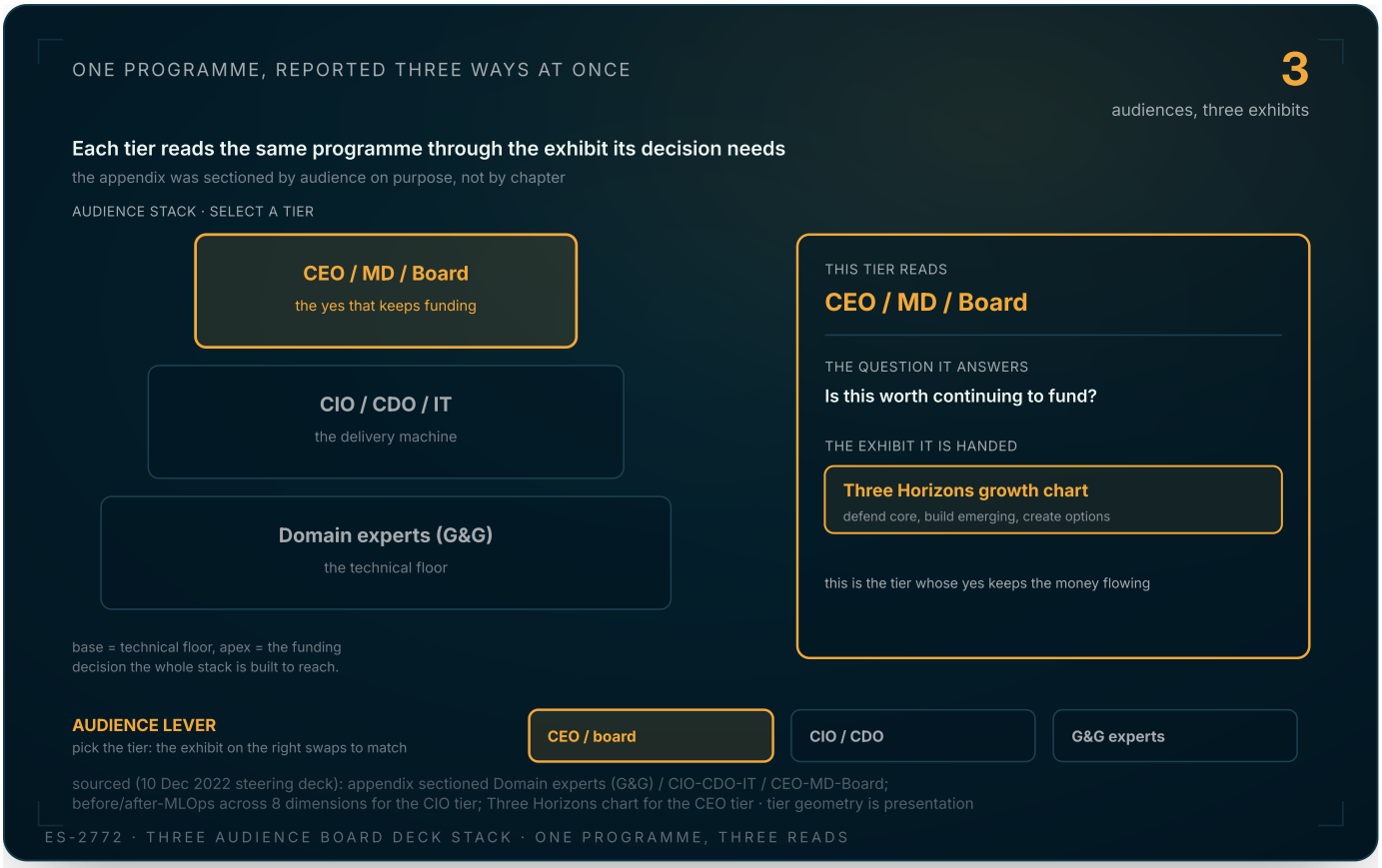
A research programme that reports itself one way reports itself to nobody in particular. The people who decide a multi-year programme's fate do not share a question. A structural geologist wants to know whether the picks are right. A CIO wants to know whether the delivery machine is maturing. A CEO wants to know whether the whole thing is still worth funding against everything else the budget could do. Hand all three the same slide and at least two of them are reading past it. The December 2022 steering deck did the opposite: its appendix was sectioned by audience on purpose, three tiers reading the same programme through three different exhibits.

The base tier was the domain experts in geology and geophysics. Their exhibit was the evidence a geoscientist actually trusts: the model's picks laid against geoscientist-validated ground truth, sinusoid-count parity in a given depth window, the dip and azimuth errors, the intervals where the model flagged a probable feature the manual pass had missed. This tier gates everything above it. If the picks do not hold up here, no amount of maturity narrative or growth framing upstairs is worth presenting.

The middle tier was the CIO, CDO and IT leadership, and their exhibit borrowed a maturity frame from McKinsey's "scaling AI like a tech native" work: a before-and-after-MLOps read across eight delivery dimensions, spanning data management, AI development, deployment, live operations, the technology stack, governance and risk, asset management, and people. The question this tier is answering is not "are the picks right" but

"is this becoming a repeatable capability or is it still a heroic one-off." A related touch in the same deck made the honesty of that frame concrete: parts of the operator's own data estate diagram were greyed out, with the legend stating plainly that greyed-out meant an area the project team was not aware of at commencement. That is the maturity story told against itself, an admission that the map was incomplete at the start, which is exactly the kind of thing a CDO trusts more than a clean diagram.

The apex tier was the CEO, managing director and board, and their exhibit was a McKinsey Three Horizons growth chart: defend the core, build the emerging, create the options. Their question is the funding question, and it is the only one that keeps the programme alive month to month. Everything below exists to earn the right to put a credible Horizon story in front of this tier.



The reporting architecture of the 10 December 2022 steering deck, drawn as three stacked audience tiers reading the same programme three different ways. The base tier is the domain experts in geology and geophysics, handed the model-versus-ground-truth evidence: do the picks hold up. The middle tier is the CIO, CDO and IT leadership, handed a before-and-after-MLOps read across eight delivery dimensions: is the delivery machine maturing. The apex tier is the CEO, managing director and board, handed a Three Horizons growth chart, defend core, build emerging, create options: is this worth continuing to fund. The audience lever raises one tier and swaps the right panel to that tier's audience, question and exhibit. The orange element is the only one that argues: the apex CEO and board tier, the one whose yes keeps the money flowing, and therefore the tier the whole architecture is built to reach. The three-audience taxonomy, the eight before-and-after-MLOps dimensions and the Three Horizons framing are sourced from the steering deck; the tier-stack geometry is presentation, not a metric.

The discipline in that layering is easy to underrate. It is tempting to write one deck and let each reader find their own slide, and it is tempting to flatten the three questions into a single "how's it going" narrative. Both fail the same way: they optimise for the writer's convenience over the reader's decision. A three-audience deck costs more to produce and it is why the reporting worked. Each tier saw the exhibit its decision needed, in the language that decision is made in, and the programme stayed legible to all three at once rather than to whichever one the author had in mind that month. We have written elsewhere in this record about the McKinsey phase-ladder operating model and the Three Horizons growth path this programme mapped itself onto; that phased operating-model piece owns the

horizon mechanics, and we do not re-derive them here. The point for this paper is narrower and about form, not content: the same status was told three ways, deliberately, and that is a governance choice.

The maturity story told against itself

The middle reporting tier deserves a closer look, because it is where a research programme is most tempted to flatter itself. "Is the delivery machine maturing" is a question with an obvious wrong answer, which is to draw a clean before-and-after diagram where everything moved from red to green. The eight-dimension before-and-after-MLOps frame in the December 2022 deck did something more useful than that, and it did it with two devices worth naming because they are transferable.

The first was the innovation funnel. Rather than reporting the delivery machine as an abstract maturity score, the deck laid out the actual sequence of things the programme had built across the year: a not-a-number remover for the dirty raw logs, a sinusoid picker, clustering approaches that were tried and set aside, self-supervised experiments, generative approaches, the path-opening morphology that recovered sinusoids where line-fitting failed, the vug detector, successive versions of the detection transformer, and the well-to-well correlation algorithm. It also tracked the data-management dashboard itself through numbered versions across the year. A funnel of this kind is honest in a way a maturity score is not, because it shows the dead ends alongside the survivors. Clustering was tried and it did not separate the sinusoids; that is in the record, not airbrushed out. A CDO reading a funnel that shows only successes distrusts it correctly. A funnel that shows the discards is one they can believe.

The second device was more pointed. The deck carried a diagram of the operator's own data estate, and parts of it were greyed out, with the legend stating in plain words that greyed-out meant an area the project team was not aware of at commencement. Read that again, because it is unusual. A vendor was telling a client's board, in a formal steering deck, that its own initial map of the client's data was incomplete, and marking exactly where. The instinct in most reporting is to hide the gaps in your own understanding, because they look like weakness. Marking them does the opposite. It tells a technically literate reader that the rest of the diagram, the parts that are not greyed out, can be trusted, precisely because the

author was willing to admit where the map ran out. A programme that will show a board the edges of its own knowledge is one whose confident claims a board can weight more heavily, not less.

Both devices point at the same governance principle, and it is the one that makes the middle tier trustworthy: a maturity report is credible in proportion to what it is willing to show going wrong. The funnel shows the discarded approaches. The greyed-out estate shows the incomplete map. Neither is a confession dragged out of the team; both are volunteered, and that is what makes the confident parts of the same deck land.

The honest programme names its own worst risk

The failure mode of research reporting is not lying. It is selective truth, the slow drift toward reporting only the dimensions that are going well. A programme that only ever shows its best metric is not governable, because the board cannot see the thing that will actually kill it coming. The most credible move in the December 2022 deck was therefore not a result. It was a risk, stated plainly: data collection was moving at a really slow pace, and that slow pace, not any modelling difficulty, was named the top programme risk.

That naming is worth dwelling on because it inverts the usual story about AI research. The romantic version of the bottleneck is always the model: the architecture, the training run, the clever loss. On this programme the binding constraint was upstream of all of that. A supervised detector can only learn from wells that actually arrive, labelled, in usable condition. When the wells arrive slowly, every downstream date slips, and no amount of model work buys the schedule back. Calling slow data the number-one risk is calling the real constraint by its name instead of the flattering one.

The mitigation was not a promise to try harder. It was a hard floor. Phase 3 was gated on a minimum of 10 to 15 wells with usable data, a number stated as a requirement rather than a hope. We can write that floor as a plain inequality on the count of usable wells:

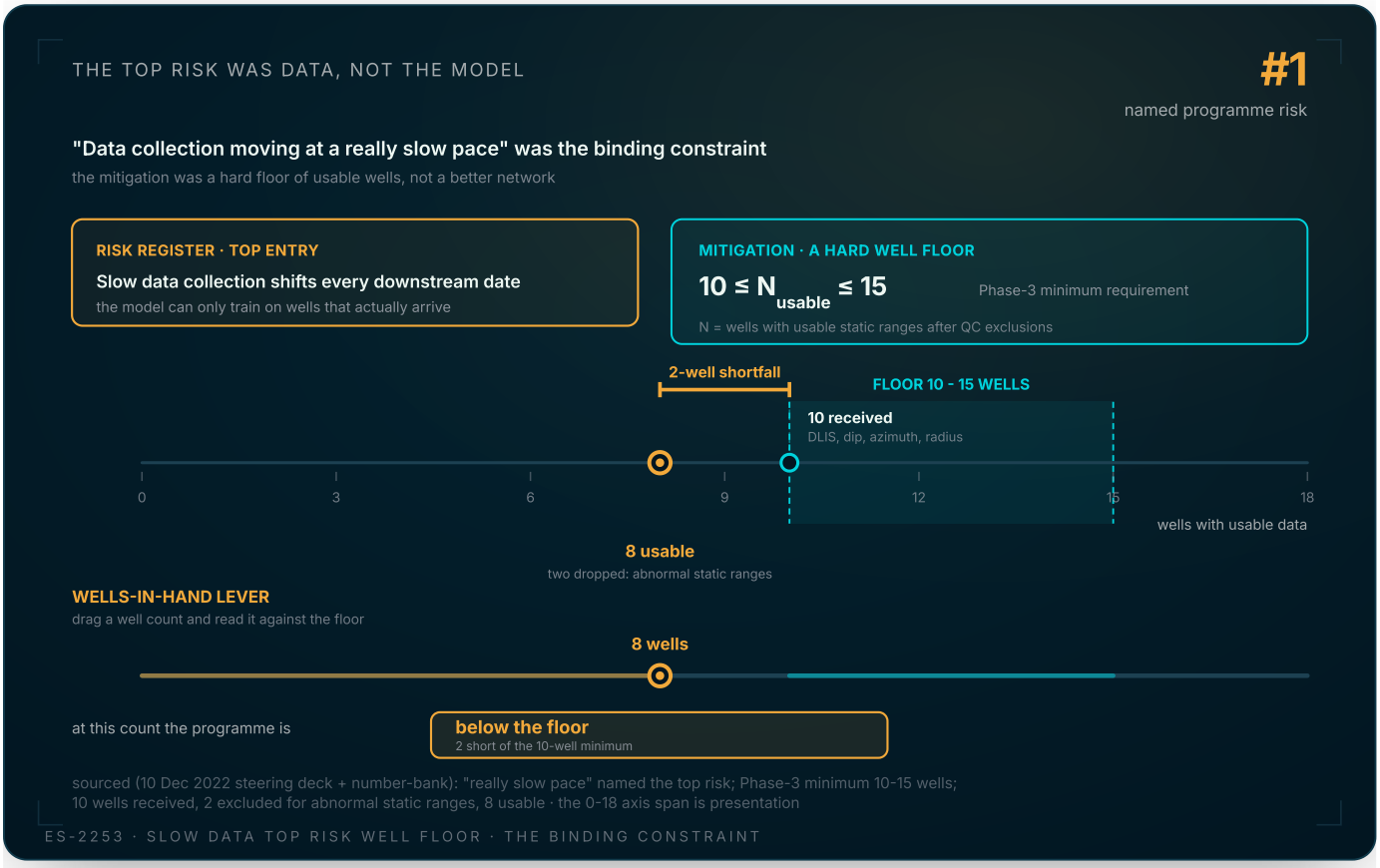
THE PHASE-3 USABLE-WELL FLOOR, STATED AS A GATE

Formula | LaTeX

Copy LaTeX

$$10 \leq N_{\text{usable}} \leq 15$$

where N-usable is the count of wells that clear quality control, not the count that were shipped. The distinction between shipped and usable is where the risk actually bit. Ten wells were received, complete with the digital log format, apparent dip and azimuth, well radius and the interpreted PDF. But two of those ten were excluded before training because their static-image value ranges fell wildly outside the normal band and defeated normalisation. That left eight usable wells against a floor whose low end was ten. The programme sat two wells below its own stated minimum, and the governance value is that this shortfall was visible in the numbers rather than hidden behind an averaged headline.



The top named programme risk plotted against the hard floor set to contain it. A single axis counts wells with usable data. The teal band marks the sourced Phase-3 minimum, ten to fifteen wells. Above the axis, ten wells received; below it, eight usable after two were excluded for abnormal static value ranges. The orange element is the only one that argues: the two-well shortfall bracket from the eight usable up to the bottom of the floor, the gap the slow-data risk actually opened. The wells-in-hand lever drags a count along the axis and flips the verdict as it crosses the floor, from below the floor through at the floor to clears with margin, so the reader feels where a given data position lands. The named top risk, the ten-to-fifteen-well minimum, the ten received and the eight usable after two QC exclusions are sourced from the 10 December 2022 steering deck and the engagement number-bank; the zero-to-eighteen axis span is presentation, not a target. The gate is stated as an inequality: ten is less than or equal to N-usable is less than or equal to fifteen.

A programme that had reported only its detection F1 that month would have looked fine, because the model trained on eight wells worked. A programme that named slow data as its top risk and set a usable-well floor put the two-well shortfall on the table where a board could see it, weigh it, and decide whether to push for more data or accept the constraint. That is what a risk register is for. It is not a place to record what already went wrong; that is the Issues column. It is a place to record what will go wrong if nothing changes, early enough that the board still has choices. Naming slow data as the number-one risk, against a hard floor the delivered data barely cleared, is the single most honest exhibit in the whole governance stack, precisely because it was the least flattering.

The floor also did something subtle to the incentives of the programme. A minimum stated as a hard requirement converts a slow, ambient problem into a specific, actionable one. "We need more data" is a wish that never quite lands on anyone's desk. "We are two wells below the ten-well minimum that Phase 3 is gated on" is a number a client's own data team can act on, because it is precise about how much and precise about the consequence. This is why the mitigation was expressed as a floor rather than as a plea to accelerate. A floor gives the slow-data risk a shared target, and it gives the board a clean decision at the gate: the wells are there or they are not, and if they are not, the programme's own rules say what happens next. The discipline is in refusing to let the constraint stay vague. The two-well shortfall was uncomfortable precisely because the floor made it countable, and a countable shortfall is one a board can govern rather than only worry about.

Why discipline, not models, is what stays fundable

Put the four pieces next to each other and a shape appears. A monthly cadence writing into a RAID log makes the programme legible week to week. A data-processing agreement signed a year early makes it lawful from the first byte. Reporting layered for three audiences makes it legible to everyone whose yes it needs. And a risk register that names slow data as the top risk, against a hard well floor, makes it honest about the thing most likely to kill it. None of these is a model. All of them are the reason the models ever got the runway to become good.

This is the inversion at the heart of the programme. The instinct, in an applied-AI research team, is to believe that funding follows results, that if the science is strong enough the governance will take care of itself. On a multi-year timescale that is not how it works. Results are lumpy and late; a research programme has long stretches where the honest

monthly report is "still working on it." Across those stretches, the thing that keeps the money flowing is not a result. It is the board's continued confidence that it can see what is happening, that the programme is lawful, that the reporting is not hiding anything, and that it could stop the whole thing cleanly if it decided to. Governance is what manufactures that confidence when the science cannot, because the science is mid-experiment.

There is a version of this argument that sounds like bureaucracy, and it is worth separating the two. The scaffolding here is not process for its own sake. Every piece of it maps to a specific failure it prevents. The cadence prevents the thirty-day blind spot. The RAID log prevents the silently relitigated decision and the forgotten risk. The early DPA prevents the unlawful transfer that cannot be undone. The three-audience deck prevents the board reading past the slide meant for the CIO. The named data risk prevents the flattering-metric drift that leaves a board blindsided. Bureaucracy is process without a failure attached to it. This was the opposite: a small, specific set of controls, each earning its place against a real way the programme could have died.

The transferable lesson is not the specific cadence or the specific clause set. It is the ordering of belief. A research programme is fundable when a board can govern it, and a board can govern it when the programme is legible, lawful, honestly reported and stoppable, in that order, independent of whether the current month's model is any good. Build that scaffolding first and the science gets the years it needs. Skip it and bet everything on the models being good on the right week, and the programme dies in one of the bad weeks instead. For a multi-year research bet, discipline is not what you do around the product. For long stretches, it is the product.

Limitations

This paper reads a governance operating model from an engagement archive, and several honest limits apply. First, the cadence frequencies we plot, roughly twenty-two stand-ups, four weekly meetings and one signed report per month, are the arithmetic of a calendar month applied to a stated daily, weekly and monthly rhythm; the specific counts are our presentation of a sourced cadence, not a logged attendance record. Second, the eight-versus-ten well figures describe one intake window on this programme; the fuller dataset used across the wider research effort was larger, and the two-well shortfall against the floor is a snapshot of a moment, not a verdict on the whole programme's data position. Third, the RAID discipline and the three-audience appendix are read from steering and board decks; a

deck records the intended operating model, and we have not independently audited every stand-up and every report against it. Fourth, the argument that discipline rather than models is what keeps a programme fundable is a claim grounded in this one multi-year engagement; it is a pattern we believe generalises to regulated applied-AI research, but a single programme is an existence proof, not a controlled study. Finally, the governance value of the early data-processing agreement rests on the legal analysis owned by the earlier chapter of this record; here we treat that ordering as established and build on it rather than re-argue it.

References

McKinsey & Company, 2021. Scaling AI like a tech native: The CEO's role. The before-and-after-MLOps maturity frame across delivery dimensions that the CIO and CDO reporting tier was built on. https://www.mckinsey.com/capabilities/quantumblack/our-insights/scaling-ai-like-a-tech-native-the-ceos-role

Baghai, Coley and White, 1999. The Alchemy of Growth. The Three Horizons growth model, defend the core, build the emerging, create the options, used for the CEO and board reporting tier. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/enduring-ideas-the-three-horizons-of-growth

European Parliament and Council, 2016. Regulation (EU) 2016/679 (General Data Protection Regulation), Article 28 (Processor). The controller-processor framework the engagement's data-processing agreement was drawn against. https://eur-lex.europa.eu/eli/reg/2016/679/oj

EARTHSCAN

Running AI R&D Like a Board-Grade Program: RAID Logs, GDPR Rails, and Three-Audience Reporting

Published May 2026. Generated 2026-07-22.
https://www.earthscan.io/whitepapers/board-grade-ai-rd-raid-logs-gdpr-rails-three-audience-reporting