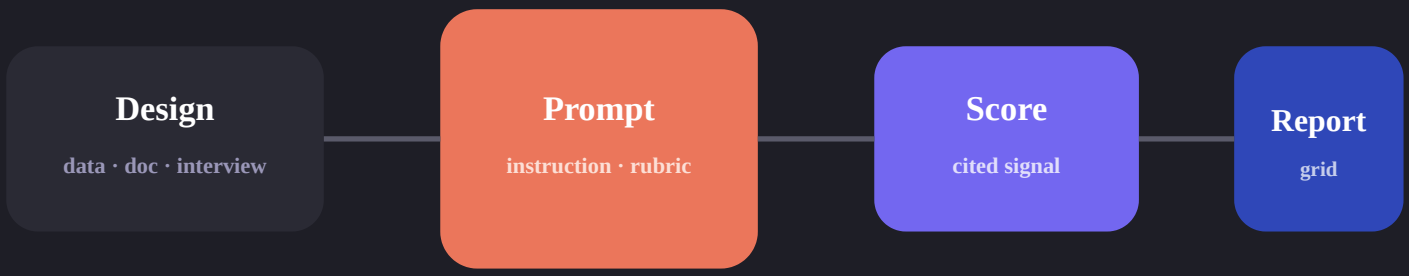
sopact

SOPACT INTELLIGENCE LIBRARY
THE PRACTICAL COMPANION



DESIGN → PROMPT → SCORE → REPORT

sopact

NO-CODE · REFERENCE-GRADE · 7 DEEP CHAPTERS

DATA DESIGN & PROMPTING.

A practitioner’s guide to designing data, documents, and interviews — and the instructions that read them — for Case, Program & Portfolio Intelligence.

THE GUIDE

Unmesh Sheth · Founder & CEO, Sopact · 2026

sopact.com

CONTENTS

A working guide, not a glossary.

Part 1 is the craft you reuse everywhere — how to design data and write the instructions that turn it into evidence. Part 2 is seven deep chapters, each written for the manager who actually runs that program: their real bind, the data to design, the prompts to write, and the reports to build.

START HERE

A survey answers a question you already knew to ask.

That is its limit. A survey is a fixed list of questions, so it can only return answers to the things you anticipated. The most valuable signal in any program — the reason a participant dropped out, the contradiction between what a grantee reported and what their narrative says, the blocker no checkbox offered — lives in the material a survey never captures: the document, the case note, the interview, the open-ended reply.

This guide is about designing for that signal. It assumes you are **not** a data scientist, do not write code, and have a program to run. What you can do — what this guide teaches — is think clearly about three things: what decision you are trying to make, what evidence would actually move it, and how to instruct a system to read that evidence consistently. Those three skills are the whole craft. Everything else is detail.

The shift this guide is built on

For two decades the bottleneck was reading. A team could collect 300 interview transcripts or 40 grantee reports, but no one had time to read them, so the qualitative material — the 95% — sat in folders while the survey's 5% went into the report. AI removed that bottleneck. A well-written instruction can now read every document on arrival, score it against a rubric you defined, and cite the sentence behind every score. The constraint is no longer reading. It is **knowing what to ask and how to ask for it.** That is a design skill, and it is learnable.

WHAT CHANGES FOR YOU

You stop choosing instruments by what's easy to tabulate, and start choosing by what reveals the most. The interview stops being "too hard to analyze."

WHAT STAYS THE SAME

You still set the rubric, the criteria, the weights. The machine reads; you decide. Judgment doesn't move to the model — it gets better evidence.

WHAT THIS IS NOT

Not a prompt-tricks list. The skill is design: framing the decision, picking the instrument, and anchoring the rubric. Tactics follow from those.

→ **This is the practical next step from each eBook**

Every chapter pairs one-to-one with a guide in the library. The eBook gives you the **why** and the strategy; this chapter gives you the **how** — the data to design, the instructions to write, the reports to build. So each chapter opens the same way: **start by downloading its eBook** at sopact.com/ebooks, then build it here.

PART ONE

The craft of data design.

Five moves you reuse in every chapter: design backward from the decision, choose the instrument that reveals the most, write the question and the instruction, build the rubric, and pick the reporting scope. Learn them once here; apply them seven times in Part 2.

TIME TO READ
~12 minutes

YOU'LL BE ABLE TO
design one instrument end-to-end

PREREQUISITE
none — no code, no stats

METHOD 1 · DESIGN BACKWARD FROM THE DECISION

Don’t ask “what should we measure?” Ask “what will we decide?”

Most measurement fails at the first step, not the last. A team opens a survey tool and starts writing questions — which means they are designing the instrument before they have named the decision it serves. Reverse it. Name the decision, name the evidence that would move it, and the instrument designs itself.

The four-line design brief

Before any instrument, write four lines. If you cannot fill them in, you are not ready to collect — and no amount of data will rescue an unframed question.

1. **Decision.** “By March we will decide whether to renew, redesign, or close the mentoring track.”
2. **Questions.** “Did mentees actually change a behavior? Which ones didn’t, and what stopped them?”
3. **Evidence.** “A specific example of the behavior at 90 days, plus the mentor’s observation of the same person.”
4. **Bar.** “Enough to defend the renewal to a skeptical board member who reads page 7.”

The two tests that kill bad questions

THE “SO WHAT” TEST

For each question, ask: if the answer is high vs low, do we do anything different? If both answers lead to the same next step, the question is decoration. Cut it.

THE OUTPUT-VS-OUTCOME TEST

“How many attended?” is an output — it proves you ran the program. “Who can now do the thing, and who can’t?” is an outcome — it proves the program worked. Funders increasingly fund the second.

! **Write the report’s headline first**

Draft the sentence you hope to write in the funder report — “68% applied the skill by day 30; the rest named transport and scheduling as blockers.” Now design the data that would let you write it honestly. If you can’t, you found the gap before the cohort, not after.

METHOD 2 · CHOOSE THE INSTRUMENT THAT REVEALS THE MOST

Four instruments. Most teams only ever reach for one.

A survey is the default because it's familiar and tabulates cleanly. But each of the four instruments answers a different kind of question, and the deepest evidence usually sits in the three a survey-only team never touches. Match the instrument to the question — and mix them on one record.

INSTRUMENT	ANSWERS THE QUESTION	DESIGN IT TO CAPTURE	FAILURE MODE
DATA	"How much, how many, how often?" — the comparable number.	Attendance, a test score, wage, cost — one clean field you can trend.	It tells you what, never why. Strand it and you're guessing at cause.
DOCUMENT	"What does the evidence behind the claim actually say?"	Essays, reports, policies, financials, site-visit notes — collect the file, not a summary of it.	It goes unread. Without an instruction to read it, it's storage.
SURVEY (MIXED)	"Where does the whole group stand, and why?"	One closed metric + one open companion ("why that number?") on the same form.	Closed-only returns the 5%. The open companion is non-negotiable.
INTERVIEW / VOICE	"What don't we know to ask?" — the blocker, the story, the surprise.	A transcript or audio reflection, lightly structured, tied to one ID.	Historically "too hard to analyze." AI themes it at scale now.

★ **The mixed-method default**
Use a **number to compare** and **words to explain**. One closed metric plus one open companion on the same record beats a 30-item Likert grid — because the words tell you what to do about the number, and the number lets you rank.

⚠ **The depth ladder**
For the same question, depth rises data → document → survey-mixed → interview — and so does effort to analyze by hand. AI flattens the effort curve, so you can choose depth by what the decision needs, not by what's easy to count.

METHOD 3 · WRITE THE QUESTION, THEN THE INSTRUCTION

Two pieces of writing decide everything that follows.

First the **question** you put to the human — phrased to surface evidence, not a polite yes. Then the **instruction** you put to the system — plain-language guidance for how to read each answer. Get both right and analysis is a formality; get either wrong and no model can save it.

1 · The question (to the human)

WEAK — INVITES AGREEMENT

“Did the program help you?” · “Are you more confident?” · “Rate your satisfaction 1–5.”

STRONG — ASKS FOR THE INSTANCE

“Describe one time in the last month you used [skill]. What happened?” · “What is the one thing still in your way?” · “Rate 1–5 — then tell us why you chose that number.”

Three rules: **be specific** (a time, a place, a number), **pair number + words**, and **leave one question fully open** so a respondent can name what you didn’t anticipate.

2 · The instruction (to the system)

ANATOMY OF A RELIABLE INSTRUCTION

```
# role & source
Read this intake essay for one applicant.
# task & scale
Score “financial-stability readiness” 0–3 using the rubric.
# evidence
Quote the exact sentence that justifies the score.
# guardrail
No evidence → answer “insufficient”. Never infer or guess.
# consistency
The same essay must always receive the same score.
```

Five parts, every time: **source, task+scale, evidence, guardrail, consistency.** The guardrail line — “say insufficient rather than guess” — is what turns a confident hallucination into a defensible result.

✓ **Write it like a brief to a careful new analyst**

You are not “prompt engineering.” You are briefing a thorough colleague who will read 400 essays exactly the way you describe. If your instruction would leave a smart intern unsure how to score a borderline case, it will leave the system unsure too. Add the missing rule.

METHOD 4 · BUILD THE RUBRIC & THE TRANSFORM

A rubric is how a feeling becomes a comparable number.

The rubric is the most important artifact you will build, and the one most teams skip. It is a short ladder that says, in observable terms, what each score means. It is what lets two readers — or a human and a machine — reach the same conclusion, and it is the only thing that makes 5 responses and 500 responses comparable.

Anchor each level to something observable

RUBRIC · "READINESS TO APPLY THE SKILL"

- 0 **No evidence.** The text doesn't speak to it at all.
- 1 **Aspiration only.** Wants to, names no concrete step.
- 2 **Concrete plan.** Names a specific action they will take.
- 3 **Demonstrated.** Gives a real example of having done it.

Test of a good anchor: could two strangers, given the same text, land on the same level? If a level needs a judgment call ("seems motivated"), rewrite it around something you could point to in the text.

Three rubric mistakes to avoid

ADJECTIVE LADDERS

"Weak / fair / strong / excellent" means twelve things to twelve readers. Replace each rung with a behavior or a fact.

TOO MANY LEVELS

0–3 or 0–4 is plenty. A 1–10 scale just relocates the ambiguity into the gaps between numbers.

NO "INSUFFICIENT" LANE

Without an explicit "not enough evidence" option, every blank gets forced onto the ladder, and your averages quietly lie.

THE TRANSFORM, IN PLAIN TERMS

Raw input
essay · transcript · PDF

→

Instruction + rubric read it
level + the cited sentence

→

A comparable signal
a score, on the record, defensible

METHOD 5 · SCOPES & REPORTS

Four zoom levels. Name the one your question actually needs.

Every report is built from four scopes. Confusion about which one you're in is why teams produce a 40-page deck that answers nobody's question. Pick the scope first; the report follows.

THE ANSWER · CELL
One response read against the rubric — a single essay scored 2/3, cited. The atom everything else is built from.

THE PERSON · ROW
Everything about one participant on one record — intake, surveys, notes, transcript. A one-page brief, not five open tabs.

THE QUESTION · COLUMN
One question across everyone — themes across 500 open-ends, pre→post movement, a retention curve, a reviewer-drift check.

THE PICTURE · GRID
All people, all programs — the funder report, the board view, aligned to whatever framework is required, every figure cited.

REPORT TYPES THE GRID SHIPS

Impact	outcomes vs commitments, cited to source
Pre / post	movement on the same people, by segment
Missing-data	who hasn't submitted — on the due date
Out-of-ordinary	the outlier worth a phone call
Multivariate	does score go with confidence? — bound at collection
Aggregate	cohort → portfolio, source one click away

MATCH THE VISUAL TO THE SCOPE

Movement (column) → a before/after bar per segment.

Pattern across people (column) → a ranked theme chart with counts.

One person's journey (row) → a timeline of touchpoints.

The whole book (grid) → a filtered table + one headline number, every cell traceable.

PART TWO

Seven programs, seven ways to think.

Each chapter takes one program manager’s real bind and works it end to end — how they think today, the data to design, the instructions to write, the rubric, the four scopes with worked numbers, the reports to build, and the mistakes to avoid. Read the one that matches your work; the method underneath is the same.

PER CHAPTER
mindset → data → prompt → rubric → scopes → report

BUILT TO
copy, adapt, and run

CHAPTER ONE · CASE INTELLIGENCE

The frontline program: when the answer is in the case note.

A case manager already holds the evidence — in notes, intake forms, and the things clients say between the survey questions. The job of data design here is to make that evidence readable in time to act on it, client by client, this week.

PAIRS WITH
Beyond Case Management

THE MANAGER
runs a service, not a spreadsheet

THE DECISION
who needs help, now

START HERE

Read **Beyond Case Management** first for the strategy, then build it here — `sopact.com/ebooks/case-intelligence`

CASE INTELLIGENCE · THE MINDSET SHIFT

Priya doesn’t need another dashboard. She needs to know who’s slipping.

Priya runs a youth-services program: 1,500 young people, 30 caseworkers, three sites. Every quarter she sends a well-being survey and gets a tidy average back — 3.9 out of 5, flat. The number is useless to her, because her real question isn’t “how is the cohort?” It’s “which eleven kids are about to disengage, and what is actually going on with them?”

> *“The survey says everyone’s fine. But I read one caseworker’s note last week and a kid had stopped sleeping and missed two sessions. That never shows up in a 4.0 average — and by the time it shows up in attendance, we’ve lost them.”*
>
> PRIYA · PROGRAM DIRECTOR, YOUTH SERVICES

Where she’s stuck today

The signal she needs is already being written — caseworkers log notes after every contact. But those notes sit in a free-text field no one reads across cases. The survey is the only thing that “rolls up,” so the program is steered by the 5% that’s easy to total, while the 95% that explains it stays invisible until a crisis.

What intelligence changes

Stop treating the case note as documentation and start treating it as **data**. Read every note on arrival against a simple risk rubric, surface the divergence between what a client *marked* and what their caseworker *wrote*, and hand Priya a ranked “call these today” list — with the exact line that triggered each flag. Same notes, read in time to matter.

→ **The reframe for this chapter**
Case intelligence is not a better survey. It is the discipline of reading the qualitative record — notes, intake, reflections — on one client ID, continuously, so the question shifts from “how did the cohort do last quarter?” to “who needs me this week, and why?”

CASE INTELLIGENCE · STEP 1

Design the data around the client, not the cohort.

Start from Priya's decision — "who do I intervene with this week?" — and work backward to the evidence. The instruments she needs aren't exotic; they're the things her team already produces, redesigned so they live on one client record and can be read together.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
DATA **Attendance & intake**	Sessions attended, demographics, goals set — on a persistent client ID minted at intake.	"Has contact dropped vs this client's own baseline?" (not vs the average)
SURVEY **Well-being pulse**	A 1–5 rating **plus** "what's the one thing hardest right now?" — short, monthly.	"Does the number agree with the sentence beside it?"
DOCUMENT **Case note**	The caseworker's free-text note after each contact — collected as text, not a checkbox.	"What did the worker observe that no form asked about?"
INTERVIEW **Reflection**	A short recorded or written reflection at 3 and 6 months, in the client's words.	"In their own telling, what changed — and what's still blocking them?"

The one design decision that unlocks the rest

Mint a **persistent client ID at intake** and attach every later instrument to it. Without it, the pulse, the note, and the reflection are three strangers; with it, they're one person's story — and the survey-vs-note contradiction becomes visible.

What *not* to do

Don't replace the note with a 12-field structured form to "make it analyzable." You'll destroy the very nuance that holds the signal. Keep the note free-text; let the instruction do the structuring.

CASE INTELLIGENCE · STEP 2

Three instructions turn notes into a “call these today” list.

Priya writes three instructions — one to read each note, one to reconcile the note against the pulse, one to find patterns across the caseload. None require code. Each names the source, the task, the rubric, and the guardrail.

INSTRUCTION 1 · READ THE NOTE (CELL)

Read each case note from the last 30 days. Detect any of: sleep, safety, housing, attendance, a new barrier. Assign an engagement flag green / yellow / red using the rubric. Quote the line that drives the flag. If the note is routine, return “no signal” — don’t inflate.

INSTRUCTION 2 · RECONCILE (ROW)

For each client, compare this month’s pulse rating with the case note. Flag a contradiction where the rating is ≥4 but the note describes withdrawal, distress, or a missed session. Name both sources.

RUBRIC · ENGAGEMENT FLAG

- G — **On track.** Attending; no new barrier named in the note.
- Y — **Watch.** A barrier appears, *or* pulse and note disagree.
- R — **At risk.** Withdrawal, a safety/attendance signal, or a plateau across two contacts.

INSTRUCTION 3 · PATTERN (COLUMN)

Across all red + yellow clients this month, cluster the named barriers into themes and count each. Surface any theme rising vs last month.

✓ **Why “quote the line” is the whole game**

The citation is what lets a caseworker trust the flag in five seconds — they see the sentence, not a black-box score — and what lets Priya defend an intervention later. Reproducibility plus a citation is the difference between “the AI said so” and “here’s the evidence.”

CASE INTELLIGENCE · STEP 3

The same client record, at four zoom levels.

Here is one month on Priya’s caseload, worked through the four scopes — with the kind of output each actually produces.

CELL · ONE NOTE

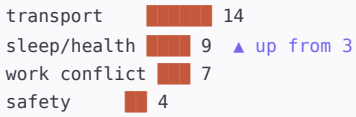
```
client P-0418 · note 2026-02-11
flag: YELLOW
signal: sleep disturbance (new)
quote: “mentioned not sleeping… did the practice but withdrawn.”
```

ROW · ONE CLIENT

P-0418 across 6 notes, 4 pulses, intake, and a 3-month reflection — one brief. The story: confidence rating rose 2→3.5, but two notes name sleep and a missed session. **Net: watch, not crisis — book a check-in.**

COLUMN · ONE QUESTION, ALL CLIENTS

Barriers across 47 watch/at-risk clients this month:



The sleep cluster tripling is the kind of thing no average surfaces.

GRID · THE WHOLE CASELOAD

A risk board — 1,500 clients by site and caseworker, sortable to **“red, contacted < 0”**: 11 names, each one click from the note behind the flag. Plus outcome movement by cohort for the board.

The insight Priya could never get from a survey

The client who rates a 4 but reads as withdrawn — **P-0512** — surfaces only because the pulse and the note sit on one record and an instruction was told to flag the disagreement. That single reconciliation is the highest-value report in the chapter.

CASE INTELLIGENCE · STEP 4

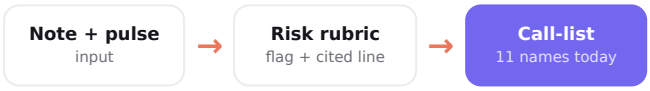
From the weekly call-list to the board’s outcome report.

The same connected record drives both the operational view (who to call) and the accountability view (did we move outcomes) — so Priya never reconciles two systems to answer one board question.

REPORTS THIS RECORD SHIPS

Call-list	ranked red/yellow clients, with the cited line + talking point
Caseworker view	my clients, my flags, this week — private to each worker
Outcome movement	pre→post on the program’s own measures, by site & cohort
Equity cut	are outcomes even across demographic groups? — filtered
Missing-data	clients with no contact in 30 days — flagged automatically

THE VISUAL THAT LANDS



Σ

Aggregate · TOC → outcomes + cost + audit

Theory of change (what the program intends to change) → reported outcome movement on the same clients + program cost-per-outcome from accounting → an audit-ready report where every figure clicks back to a note. Outputs (sessions delivered) become outcomes (who changed) become evidence (cited).

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Averaging away the person. Reporting only cohort means is how you miss the 11 who need you. Keep the row scope alive.

2 · Killing the note. Forcing free-text into dropdowns to “analyze” it destroys the signal. Read the text instead.

3 · No baseline per client. “Attendance dropped” is meaningless without *this* client’s own starting point. Trend each person against themselves.

CHAPTER TWO · APPLICATION INTELLIGENCE

Selection at scale, without the committee running on vibes.

Eight hundred applications, forty volunteer reviewers, three weeks to decide. The design challenge is to read every application against the same rubric, surface where scores drift, and keep a citation behind every judgment — so the decision is fair and defensible.

THE MANAGER
a review lead under deadline

THE DECISION
who advances to the shortlist

THE RISK
bias, drift, a yes you can't defend

START HERE

Read **Application Intelligence** first for the strategy, then build it here — `sopact.com/ebooks/application-intelligence`

APPLICATION INTELLIGENCE · THE MINDSET SHIFT

Marcus can’t read 800 essays deeply. So the committee reads the famous names.

Marcus runs a fellowship: 800 applications, 40 volunteer reviewers, a three-week window. Each reviewer gets a stack, scores on a 1–5 form, and the committee meets to decide. By hour three nobody is reading the way they read at hour one — and the applications with a recognizable employer or university quietly float up. Marcus knows the process isn’t fair; he just can’t prove where it breaks.

> *“Two reviewers gave the same essay a 2 and a 5. The committee averaged it to a 3.5 and moved on. I have no idea which reviewer actually read it — and neither does the board if they ever ask.”*
>
> MARCUS · FELLOWSHIP REVIEW LEAD

Where he’s stuck today

The scoring form records a number but not whether the essay was read, or why. Reviewer drift only appears in the export after the committee has met — too late to fix. And the most decision-relevant evidence (a concrete example of initiative buried on page 2) is exactly what a skimming reviewer misses.

What intelligence changes

Read every application **on arrival** against the rubric, with identity masked, and attach the sentence behind each score. The committee opens to a ranked shortlist with citations — not a queue — and reviewer outliers are flagged *before* the meeting, while there’s time to calibrate. The human still decides; the reading and first-pass scoring stop being a lottery.

→ **The reframe for this chapter**
Application intelligence is not “AI picks the winners.” It is making the fair process — blind, rubric-anchored, calibrated, cited — the **default path** instead of the exception a tired committee can’t hold to.

APPLICATION INTELLIGENCE · STEP 1

Design the form so the evidence is there to read.

A fair review starts before the deadline, in how the application is built. Ask for the example, not the adjective; collect what can be masked; and bind every artifact to one applicant ID so the essay, the recommendation, and the later outcome are one record.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
DOCUMENT **Essay / statement**	An open prompt that asks for a specific instance: “describe a time you…” — not “why are you a leader?”	“Is there a real example here, or only a claim?”
DOCUMENT **Recommendation**	The referee’s narrative — collected as text, with named behaviors prompted.	“Does the referee corroborate the applicant’s own claim?”
SURVEY **Structured intake**	Eligibility + demographics, stored separately so they can be **masked** at review.	“Can we review blind, and still cut the shortlist by equity later?”
DATA **Work sample / portfolio**	A link or file scored on the same rubric as the essay.	“Does the work back up the words?”

The design choice that designs bias out
Store identity fields (name, school, employer, photo) **separately** from the material reviewers read. Blind-by-default isn’t a setting you remember to switch on — it’s how the data is structured, so anchoring and prestige effects lose their hook.

What *not* to do
Don’t ask “rate your own leadership 1–5.” Self-ratings compress and overclaim. Ask for the instance and let the rubric rate it — the example is the evidence, the self-rating is noise.

APPLICATION INTELLIGENCE · STEP 2

Score blind, cite the line, flag the reviewer who drifts.

Three instructions: one reads each masked essay against the rubric, one checks whether the referee corroborates the claim, one finds reviewers scoring far from the panel. Each ends with the same guardrail — cite, or say there isn't evidence.

INSTRUCTION 1 · SCORE THE ESSAY (CELL)

Read this essay with name, school, employer hidden. Score each criterion 0–3 on the rubric. Quote the sentence behind each score. Any criterion with no evidence → "insufficient." Same essay, same score, every time.

INSTRUCTION 2 · CORROBORATE (ROW)

Compare the essay's initiative claim with the recommendation. Mark corroborated / silent / contradicted; quote the referee line.

RUBRIC · "DEMONSTRATED INITIATIVE"

- 0 **No evidence** in the text.
- 1 **Claims it**, gives no example.
- 2 **One concrete example.**
- 3 **Sustained example with a result.**

INSTRUCTION 3 · REVIEWER DRIFT (COLUMN)

For each reviewer, compute the gap between their scores and the panel median on shared applications. Flag anyone consistently >1 point off.

✓ **The committee's job changes — for the better**

Reviewers stop being score-generators and become judges of evidence. They open each applicant to a cited summary and a draft score, and spend their scarce attention on the borderline cases and the flagged disagreements — not on reading 800 cold.

APPLICATION INTELLIGENCE · STEP 3

From one essay to a defensible shortlist.

One fellowship round of 800, worked through the four scopes.

CELL · ONE ESSAY

```
applicant A-3312 (masked)
initiative: 2 / 3
quote: “I started a Saturday clinic and ran it
for a year…”
communication: insufficient — no evidence in text
```

COLUMN · REVIEWER DRIFT

```
vs panel median (shared apps):
R-07  +0.2  ok
R-19  +1.4  scores high · flag
R-22  −1.1  scores low · flag
```

Caught before the committee meets, not in the post-mortem.

ROW · ONE APPLICANT

A-3312 — essay, recommendation, work sample on one brief. Initiative 2, corroborated by the referee (“built it from nothing”); leadership 3; communication thin. **Net: advance, flag communication for interview.**

GRID · THE ROUND

A ranked shortlist with every score cited, plus the **fairness view**: advance-rate by applicant segment, time-to-decision, and the 6 disagreements worth a second read. Board-defensible as one export.

> ★ **The thing a scoring form can never give you**
> “Show me every declined applicant who scored 3 on initiative but was marked down on a criterion with no evidence” — a one-query audit of your own fairness. The form gave you a number; the record gives you accountability.

APPLICATION INTELLIGENCE · STEP 4

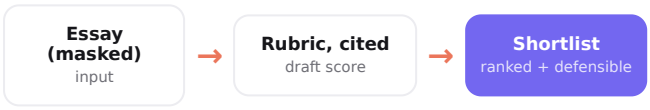
The decision record that survives the board’s question.

Selection is contested — every yes is a no to someone else. The report that matters is the one that lets you explain any decision, months later, with the evidence intact.

REPORTS THIS RECORD SHIPS

Shortlist	ranked, every score cited to a sentence
Decision record	per applicant: scores, citations, reviewer notes — funded & declined
Fairness view	advance-rate & score spread by segment
Reviewer report	drift, load, and calibration per reviewer
Missing	applications short a document; reviewers behind

THE VISUAL THAT LANDS



Σ Aggregate · the applicant becomes the participant

The same ID carries past the decision: the applicant you scored is the grantee or scholar you track for years. Selection rubric → onboarding baseline → outcome — so you can finally ask “did our highest-scored applicants actually do best?” and improve the rubric from real results.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Prestige proxies. Scoring “quality of writing” rewards privilege, not potential. Anchor on the example, not the polish.

2 · Letting the rubric drift. If reviewers reinterpret “initiative” privately, you have 40 rubrics. One anchored rubric, applied by the read, holds.

3 · Scores without citations. A number you can’t trace is indefensible the moment a declined applicant appeals.

CHAPTER THREE · GRANT INTELLIGENCE

Twenty-eight grantees, twenty-eight spreadsheets, one board meeting.

The work isn’t the award — it’s the years after. Quarterly metrics in a different format every time, narratives no one reads in sequence, and a three-week scramble to assemble the board packet. The design job is to make each report read itself and roll up on one grantee record.

THE MANAGER
a foundation program officer

THE DECISION
who needs a check-in; is the portfolio on track

THE PAIN
reconciliation, every cycle

START HERE

Read **Grant Intelligence** first for the strategy, then build it here — `sopact.com/ebooks/grant-intelligence`

GRANT INTELLIGENCE · THE MINDSET SHIFT

Dana spends three weeks assembling what the data already knows.

Dana is a program officer managing 28 active grants across a health portfolio. Every quarter, grantees send reports — some PDFs, some spreadsheets, some email narratives, each with its own definition of “served.” Before each board meeting she spends three weeks reconciling them into one deck, and the qualitative insight — the reason a grantee is behind — gets lost on page 7 of a report no one finished.

> *“A grantee reported ‘1,200 served.’ Their narrative said a clinic closed for two months. Both were true. Nobody connected them until I happened to read the whole thing — at 11pm, the night before the board.”*
>
> DANA · FOUNDATION PROGRAM OFFICER

Where she’s stuck today

Each grantee is a separate format on a separate thread. The metric and the story that explains it live in different places, so a number can look fine while the narrative quietly contradicts it. And because nothing carries forward, year 1’s report can’t speak to year 3.

What intelligence changes

One form per cycle on a persistent grantee ID; the AI reads the narrative *against* the reported KPIs and flags the disagreements; the board packet pulls automatically — same data the program officer sees, with the context attached. Three weeks of assembly becomes a query.

→ **The reframe for this chapter**
Grant intelligence treats the award as the **handoff, not the finish line.** The grantee’s proposed metrics become the indicators you track; their record grows each cycle; and the board sees movement, not a freshly-assembled snapshot.

GRANT INTELLIGENCE · STEP 1

One form, one ID, a schema that can evolve.

The design goal is comparability without rigidity: every grantee reports on the same structure, but the structure can grow across years without breaking the history. Collect the number and the narrative on the same form — never on two threads.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
DATA **Process KPIs**	Unique served, services delivered, events — one shared definition in a data dictionary.	“Are two grantees counting ‘served’ the same way?”
SURVEY **Narrative prompts**	“What changed this quarter, and what got in the way?” on the same form as the KPIs.	“Does the story explain the movement in the number?”
DOCUMENT **Reports & artifacts**	The grantee’s own report, financials, logic model — uploaded, read end-to-end.	“What’s in the attachment that never reached the form?”
INTERVIEW **Site-visit note**	The officer’s mid-cycle observation, on the same grantee record.	“What did we see on the ground that the report omits?”

Additive schema, versioned indicators
When you refine a question in year 3, don’t overwrite the year-1 version — **version it.** Year-1 grantees keep their question; year-3 grantees get the new one; cross-cycle roll-ups still work because the definitions are versioned, not replaced.

What *not* to do
Don’t let each grantee invent its own template. The moment formats drift, every cycle becomes a reconciliation project. Shared form, shared definitions, grantee-specific context inside it.

GRANT INTELLIGENCE · STEP 2

Make the narrative check the number.

The highest-value instruction in grant work is reconciliation: read the story against the KPI and flag where they disagree. Add an outcome-confidence rubric so “we made progress” gets graded on the evidence attached, not the confidence of the prose.

INSTRUCTION 1 · RECONCILE (ROW)

```
Read the quarterly narrative against the reported
KPI values. Flag any disagreement (e.g., “served” up
but narrative describes a closure). Map each outcome
to its logic-model branch; cite the paragraph.
```

INSTRUCTION 2 · THEME BARRIERS (COLUMN)

```
Across all 28 grantees this quarter, cluster the
named implementation barriers and count each.
Surface any barrier shared by ≥5 grantees.
```

RUBRIC · OUTCOME CONFIDENCE

- 0 **Claimed**, no evidence attached.
- 1 **Narrative only.**
- 2 **Metric + matching narrative.**
- 3 **Metric + narrative + source document.**

INSTRUCTION 3 · PORTFOLIO VIEW (GRID)

```
Roll spend-vs-budget and outcome confidence by
program. Cite the grantee record behind every
figure.
```

✓ **The reconciliation flag is the product**

Dana’s 11pm discovery — “served up, but a clinic closed” — becomes an automatic flag the day the report lands. That single instruction converts a buried contradiction into a visible, cited signal she can act on before the board, not after.

GRANT INTELLIGENCE · STEP 3

From one report to the board view.

One quarter across a 28-grant health portfolio, worked through the four scopes.

CELL · ONE REPORT

```
grantee G-14 · Q3
KPI “served”: 1,200
confidence: 1 / 3
flag: narrative names a 2-month closure
quote: “clinic paused Aug–Sep for staffing.”
```

ROW · ONE GRANTEE

G-14 across 6 cycles — schema evolved twice, history intact. Trajectory: served volatile, outcome confidence rising 1→2; staffing the recurring barrier. **Net: check-in on staffing, not on numbers.**

COLUMN · SHARED BARRIERS

Staffing is a portfolio problem, not 17 separate ones — a funder-strategy signal.

GRID · THE PORTFOLIO

Board view: spend-vs-budget by program, outcome confidence, the staffing flag — each figure one click from a grantee report. Assembled by default, not over three weeks.

★ **The portfolio insight no single report shows**

Seventeen of 28 grantees naming staffing isn’t 17 anecdotes — it’s a finding that could redirect the foundation’s strategy toward capacity grants. Only the column scope, across all grantees, makes it visible.

GRANT INTELLIGENCE · STEP 4

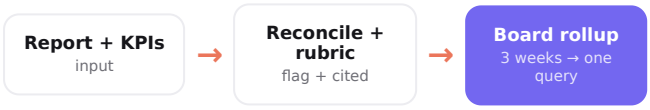
TOC, financials, and the audit trail on one record.

The reason grant work fragments is that outcomes, money, and compliance live in three systems. On one grantee record, they’re three views of the same evidence — and the board packet assembles itself.

REPORTS THIS RECORD SHIPS

Program impact	outcomes vs commitments, narrative + metrics, TOC-aligned
Check-in list	grantees flagged by reconciliation — who to call
Portfolio rollup	spend-vs-budget, outcome confidence, by program
Missing-data	who hasn’t submitted — on the due date
Board packet	multi-year, role-based, every figure cited

THE VISUAL THAT LANDS



Σ Aggregate · TOC → reported → financial → audit

Theory of change → reported quarterly metrics + narrative → financial actuals from the ERP/accounting handoff → the compliance packet — all on the same persistent grantee ID, every number traceable to its source. The comprehensive report is a join, not a rebuild.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Number without narrative. A KPI with no story can’t be interpreted — collect both on one form.

2 · Overwriting the schema. Change a definition and you break year-over-year. Version it instead.

3 · Reporting outputs as outcomes. “1,200 served” is activity. Pair it with what changed for them.

CHAPTER FOUR · ESG INTELLIGENCE

The score said 72. The workers said something else.

A provider rating is a point-in-time number built from public disclosures. The evidence that actually matters — the sustainability report no one read, the contradiction between policy and the worker survey — sits in documents and interviews. The design job is to read them on one entity record, year over year.

THE MANAGER
an ESG analyst or procurement lead

THE DECISION
commit, monitor, or flag this entity

THE BAR
prove diligence was effective, not just performed

START HERE

Read **ESG Intelligence** first for the strategy, then build it here — `sopact.com/ebooks/esg-intelligence`

ESG INTELLIGENCE · THE MINDSET SHIFT

Lena has three provider scores for one company — and they don’t agree.

Lena runs ESG diligence at an impact fund — 40 portfolio companies, plus a supplier base behind several of them. She buys ratings, but the same company scores 47, 61, and 74 across three providers, each refreshed on its own cycle from public data. When an LP asks “is your diligence actually preventing harm?” a number can’t answer. Only a record of what she saw, when, and what changed can.

> *“Governance scored fine for three quarters while a labor issue was screaming in the worker interviews — which no rating provider had read, and frankly neither had we. They were a folder, not a finding.”*
>
> LENA · ESG ANALYST, IMPACT FUND

Where she’s stuck today

Scores are point-in-time, provider-dependent, and built from what the company *says*, not what is happening. The DDQ, the sustainability PDF, the worker survey, and the audit each live in a different place, so a policy claim and the worker reality that contradicts it never meet.

What intelligence changes

Keep the score as a first-pass anchor, but read every uploaded document end-to-end against your own E/S/G rubric, flag where policy and worker voice diverge, and carry it all on **one entity ID** from intake forward — so the effectiveness chain a regulator or LP wants is assembled by default, not reconstructed at audit.

→ **The reframe for this chapter**
ESG intelligence fixes the **input**, not the model. The score becomes one signal among many; the cited, longitudinal record — commitment → evidence → corrective action → re-survey — is what actually proves diligence worked.

ESG INTELLIGENCE · STEP 1

Design two modes onto one entity record.

ESG diligence runs in two modes: a structured DDQ at intake (“what does the company say?”) and continuous evidence after (“what is actually happening?”). The design goal is to land both on the same entity ID so claims can be checked against reality across cycles.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
SURVEY **DDQ at intake**	40–150 questions across E/S/G — closed compliance checks plus open commitments.	“What did they commit to, that we can verify later?”
DOCUMENT **Sustainability report**	The full PDF, policies, certifications, audit findings — uploaded, read page by page.	“Is this substantive, or template language?”
INTERVIEW **Worker / community voice**	Anonymous worker-voice surveys with open responses, tied to the site/supplier ID.	“Does the floor match the policy?”
DATA **Corrective actions**	CAPs and re-survey results linked to the original finding.	“Did the fix actually change conditions?”

One entity ID, issued at the first DDQ
The same ID carries through every form, audit, and worker survey — surviving name changes and tier moves. Without it, “did the corrective action work?” is unanswerable; with it, the re-survey sits next to the original finding.

What *not* to do
Don’t treat the provider score as the diligence. It’s a screening anchor. The work — reading the documents, hearing the workers, tracking the fix — is what an LP and a regulator actually need.

ESG INTELLIGENCE · STEP 2

Read the PDF the score never opened.

Three instructions: read the sustainability report against your rubric, reconcile the policy claim against worker voice, and theme worker feedback across suppliers. The rubric grades evidence quality — template language scores low no matter how confident.

INSTRUCTION 1 · READ THE REPORT (CELL)

Read the sustainability report end-to-end against the E/S/G rubric. Score each pillar 0–3. Mark template vs substantive language. Cite the paragraph behind each score; missing evidence → "not evidenced."

INSTRUCTION 2 · RECONCILE (ROW)

Where a policy claim conflicts with the worker survey, flag it and quote both passages. Rank by severity.

RUBRIC · EVIDENCE QUALITY

- 0 **Template language**, no specifics.
- 1 **Policy stated**, no proof.
- 2 **Policy + audit or metric.**
- 3 **Policy + evidence + worker corroboration.**

INSTRUCTION 3 · THEME WORKER VOICE (COLUMN)

Across all supplier worker surveys this quarter, cluster risks (overtime, retaliation, safety); count by site; surface any rising vs last quarter.

✓ **Citations are non-negotiable here**

CSDDD asks for evidence that diligence is effective over time. Every score must trace to the specific passage that generated it — that traceability is exactly what satisfies a regulator and what survives an LP's challenge.

ESG INTELLIGENCE · STEP 3

From one report to a supplier portfolio.

One quarter across 200 suppliers, worked through the four scopes.

CELL · ONE REPORT

```
entity E-088 · 64-page report
social: 1 / 3 · template language
quote: "we are committed to fair labor practices"
note: no metric, no audit attached
```

ROW · ONE ENTITY

E-088 — DDQ + report + worker survey + audit on one brief. Policy says "no forced overtime"; 9 worker responses describe it. **Net: flag for corrective action; re-survey in Q3.**

COLUMN · WORKER-VOICE RISK

```
"forced overtime" mentions, Q3:
3 Vietnam sites 12% ▲ from 3%
portfolio avg    4%
linked to: same sites flagged in 2024 audit
```

A query, not a multi-week investigation across 200 surveys.

GRID · THE EFFECTIVENESS CHAIN

Per entity: commitment → evidence → corrective action → re-survey → outcome — assembled by default, CSDDD-ready, every link cited.

★ **The query a score can never answer**

"Which suppliers flagged retaliation risk in their Q3 worker survey, and did our corrective action move it?" Forced-overtime mentions falling 12% → 2% after a CAP, with the re-survey tied to the original finding — that's effectiveness, proven.

ESG INTELLIGENCE · STEP 4

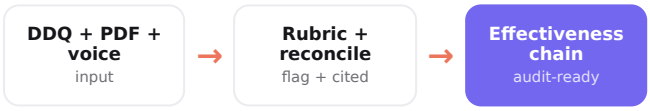
The evidence chain that’s ready before the audit.

The point of the architecture is that the report regulators and LPs want is a by-product of doing the work — not weeks of reconstruction at audit time.

REPORTS THIS RECORD SHIPS

Entity brief	DDQ + report + worker voice + audit, scored & cited
Contradiction log	policy-vs-reality flags, ranked by severity
Worker-voice themes	risk clusters across suppliers, by site, trended
CSDDD chain	commitment → evidence → CAP → re-survey, per entity
LP / board ESG	portfolio view, every figure traceable

THE VISUAL THAT LANDS



Σ Aggregate · commitment → evidence → outcome

Intake commitment → mid-cycle evidence + corrective action → re-survey vs baseline — one entity ID across all of it, so “did the diligence prevent harm?” is answered with a cited, longitudinal record. Frameworks (IFC, IRIS+, the Five Dimensions) are cited as the standard of care, not certified against.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Trusting the score. A point-in-time number can’t prove effectiveness over time. Read the documents.

2 · Scoring policy, not practice. “We are committed to…” is template language. Grade evidence, not intent.

3 · Snapshots, not chains. Without re-survey on the same ID, you can’t show a corrective action worked.

CHAPTER FIVE · IMPACT INTELLIGENCE

The LP report is three weeks away. Or overnight.

A fund holds the widest context it will ever have at due diligence — then throws most of it away. The design job is to score the impact thesis from the documents themselves and carry every finding forward, so the quarterly report writes itself from an evidence base that compounds.

THE MANAGER
an impact-fund principal

THE DECISION
invest, and how to monitor the thesis

THE FRAMEWORK
the Five Dimensions, IRIS+, scored not slided

START HERE

Read **Impact Intelligence** first for the strategy, then build it here — `sopact.com/ebooks/impact-intelligence`

IMPACT INTELLIGENCE · THE MINDSET SHIFT

Sam reads the pitch deck once, then rebuilds context from zero, every quarter.

Sam is a principal at a 25-company impact fund. Diligence is rigorous — decks, models, founder interviews, a Five Dimensions scoring at the IC. Then the findings die at the investment memo. Ninety days later the LP wants a progress update, and the team reopens folders no one has touched, re-reading what the fund already knew. The frameworks everyone references — Five Dimensions, IRIS+, theory of change — sit as PDFs, not as anything the data runs on.

> *"We score the Five Dimensions beautifully at the IC. Then nobody touches the score again until the next memo — so Contribution and Risk are basically narrative paragraphs by the time the LP reads them."*
>
> SAM · PRINCIPAL, IMPACT FUND

Where he's stuck today

Context utilization is maybe 5%. The deck is read once, the interview produces a memo, quarterly updates arrive as disconnected spreadsheets. The framework is decoration because the qualitative layer — where Dimensions 1, 4, and 5 actually live — was never processable at scale.

What intelligence changes

Score the Five Dimensions from the deck and interview at DD, on the same rubric you'll use every quarter; carry the theory of change as a living model; and let the LP report generate from the connected record. Context utilization moves toward 95%, and Contribution stops being a paragraph.

→ **The reframe for this chapter**
Start impact measurement at **due diligence, not at LP reporting.** The earliest structured record wins — every baseline commitment captured at DD becomes the reference point for the next ten quarters.

IMPACT INTELLIGENCE · STEP 1

Design the diligence to compound.

The richest context window is DD. Design it so every artifact becomes structured, queryable, and reusable — not a memo that's read once. The Five Dimensions are your scoring spine; pair every metric with the qualitative companion that explains it.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
DOCUMENT **Deck & model**	The pitch and financials — uploaded, with the implicit theory of change extracted.	"What outcome is claimed, and is it evidenced?"
INTERVIEW **Founder**	The DD interview transcript — where Contribution & Risk evidence lives.	"Is there a counterfactual, or just a story?"
DATA **IRIS+ metrics**	The selected metric set, captured at DD on a persistent investee ID.	"Are these comparable across the portfolio?"
SURVEY **Quarterly + Lean Data**	A metric + an open companion each cycle; stakeholder voice where it matters.	"Does this quarter confirm or challenge the DD thesis?"

Pair every IRIS+ metric with a narrative companion

A number without context is ambiguous. Each metric gets an open-ended "what's behind this?" on the same form — AI themes the qualitative layer across the portfolio in minutes, and the metric finally means something.

What *not* to do

Don't treat the theory of change as a slide filed after the IC. Capture it as a structured object at DD so each quarter can confirm or challenge it — a frozen ToC is compliance documentation, not a model.

IMPACT INTELLIGENCE · STEP 2

Score the dimensions from the evidence.

Three instructions: extract and score the Five Dimensions at DD, reconcile each quarter against the DD baseline, and theme stakeholder voice across the portfolio. The rubric for Contribution (D4) is where most funds wave their hands — anchor it to a counterfactual.

INSTRUCTION 1 · SCORE AT DD (CELL/ROW)

From the deck + founder interview, extract the theory of change and score the Five Dimensions 0–3. Cite every score. Where a dimension has no evidence, mark "not yet" — don't infer.

INSTRUCTION 2 · RECONCILE (ROW)

Each quarter, compare the submission to the DD baseline. Flag drift from the original thesis; re-score the dimension that changed.

RUBRIC · CONTRIBUTION (D4)

- 0 **"We contributed"** — no basis.
- 1 **Plausible story**, no comparison.
- 2 **Stakeholder attribution** captured.
- 3 **Comparison group** or counterfactual.

INSTRUCTION 3 · PORTFOLIO VOICE (COLUMN)

Across all investees, theme stakeholder responses; surface companies with declining retention and quote what founders said about it.

✓ **The query that was a four-week investigation**

"Which portfolio companies show declining customer retention, and what did founders say about it last quarter?" — two analysts, four weeks in 2022; one query now, because the metric and the narrative sit on the same investee record.

IMPACT INTELLIGENCE · STEP 3

From one claim to the LP narrative.

One quarter across a 25-investee fund, worked through the four scopes.

CELL · ONE CLAIM

```
investee I-07 · DD
D4 contribution: 1 / 3
quote: “our financing expanded access…”
note: no comparison group — story only
```

ROW · ONE INVESTEE

I-07 from DD to Q4 — one record. D1 confirmed by 84% of borrower interviews; D3 baseline-endline complete; D4 still narrative. **Net: ask for a comparison cohort next cycle.**

COLUMN · ONE DIMENSION, PORTFOLIO

```
D4 contribution evidence, 25 investees:
strong (3)   ▇▇▇ 6
partial (2)  ▇▇▇▇▇▇ 12
weak (0–1)   ▇▇▇▇ 7 · thin
```

Where the fund’s impact claim is weakest — an LP would ask exactly here.

GRID · THE FUND

Six LP-ready reports per investee — scorecard, gap memo, narrative, trend, exit, cross-portfolio — generated overnight from the connected record, every figure cited.

★ **The view an LP actually wants**

Not a glossy number — a portfolio map of where the evidence is strong and where it’s thin (7 of 25 weak on Contribution), so the fund can shore up its weakest claims before the LP finds them. That honesty is only possible when dimensions are scored, not slided.

IMPACT INTELLIGENCE · STEP 4

Generate the LP report. Don’t assemble it.

A fund running compounding measurement doesn’t rebuild the report from spreadsheets each cycle — it queries the connected record. The reading work is what most funds underestimate; the reporting is the easy part once the record exists.

REPORTS THIS RECORD SHIPS

Investee scorecard	Five Dimensions, evidence-linked, with trend
Gap & risk memo	drift from thesis, thin dimensions, flags
LP narrative	outcomes vs commitments, per LP framework
Longitudinal trend	DD → quarterly → annual, one timeline
Exit summary	entry-to-exit impact record

THE VISUAL THAT LANDS

Σ Aggregate · living ToC → IRIS+ → LP framework

A living theory of change → quarterly evidence + IRIS+ metrics with qualitative companions + cost → the LP portfolio narrative, every figure cited — and the *same* record feeds each LP’s framework, so you report many ways from one source.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Starting at LP reporting. Context stays at zero. Start at DD so it compounds.

2 · Dimensions as labels. Five headers in a report aren’t a framework. Score them, cite them.

3 · Metric without narrative. A 40-metric IRIS+ export with no story reads like a CRM dump.

CHAPTER SIX · LEARNING INTELLIGENCE

94% completion. 4.3 satisfaction. Did anyone change?

The funder doesn’t ask whether learners liked the training — they ask whether the skill reached the job at 90 days. The design job is to connect one learner across pre, mid, and post, and to ask for the example instead of the intent.

THE MANAGER
a workforce / L&D program lead

THE DECISION
renew, redesign, or coach

THE BAR
Kirkpatrick Level 3–4, not Level 1

START HERE

Read **Learning Intelligence** first for the strategy, then build it here — `sopact.com/ebooks/learning-intelligence`

LEARNING INTELLIGENCE · THE MINDSET SHIFT

Rosa collected the proof. It’s in four systems with four different IDs.

Rosa runs a 120-person workforce cohort. The LMS shows 94% completion; the post-survey shows 4.3 satisfaction. But the grant renewal asks for behavior change at 90 days — and that data exists in a follow-up spreadsheet that can’t be linked back to the pre-test or the intake. Her team spent three weeks trying to reconcile it and still couldn’t connect the records. She’s about to lose a renewal over data she actually collected.

> *“Every tool gives the same person a new ID. The LMS, the survey, the 90-day follow-up — by the time I try to connect them, the cohort has graduated and the matching is a guess.”*
>
> ROSA · WORKFORCE PROGRAM DIRECTOR

Where she’s stuck today

Completion and satisfaction (Levels 1–2) are easy because they don’t need longitudinal tracking. Behavior and results (Levels 3–4) require connecting the same learner across time — and four tools issuing four IDs make that a manual reconciliation that never finishes before the deadline.

What intelligence changes

Mint one persistent learner ID at enrollment; every instrument inherits it — pre-test, weekly pulse, post-test, mentor observation, 90-day follow-up. Ask for a specific example of using the skill, score it against a transfer rubric, and the Level 3–4 report is a default output, not a stretch goal.

→ **The reframe for this chapter**
The cascade — Reaction → Learning → Behavior → Results — only holds on one spine. Design the persistent learner ID **before the first cohort enrolls**, and Level 3–4 stop being a retrofit.

LEARNING INTELLIGENCE · STEP 1

Design pre, mid, and post as one instrument set.

The highest-leverage decision happens before the cohort opens: design the baseline, the formative checks, and the follow-up together, with matched anchors, on one learner ID. Write the 90-day question at the same time as the intake — not six weeks later when it won't match.

INSTRUMENT	DESIGN IT TO CAPTURE	THE QUESTION THAT MAKES IT USEFUL
DATA **Pre/post test**	Identical items at baseline and post, scored per individual on one ID.	"Did *this* learner's knowledge move — not the group average?"
SURVEY **Confidence open-end**	A self-rating + "tell us why" — scored separately from knowledge.	"Do they know it but won't apply it? (coach, not re-train)"
INTERVIEW **Mentor observation**	A structured rubric the mentor completes against 2–4 named behaviors.	"Does someone else see the behavior, or only the learner?"
DOCUMENT **90-day reflection**	The learner's own account of using the skill on the job.	"Is there a specific instance, or just intent?"

Name 2–4 observable behaviors first
Write them as actions ("conducts a structured weekly 1:1"), not qualities ("is a better leader"). Baseline them at enrollment and re-measure at 30/60/90 days from learner and mentor — on the same ID. The named behavior is what makes Level 3 measurable.

What *not* to do
Don't send a bulk follow-up email to whoever opens it. Response rates collapse to ~12% and nothing links back. Personalized links tied to the original record produce ~3× the response, auto-joined.

LEARNING INTELLIGENCE · STEP 2

Ask for the example. Score the transfer.

Three instructions: score transfer from the reflection (intent alone caps at 1), reconcile knowledge against confidence, and theme the blockers across the cohort. Behavioral specificity is the whole trick — “have you used it?” overclaims by two-to-four times.

INSTRUCTION 1 · SCORE TRANSFER (CELL)

From the 90-day reflection + mentor note, score “applied on the job” 0–3. Require a specific example — intent alone is a 1. Quote it. Score knowledge and confidence separately.

INSTRUCTION 2 · RECONCILE (ROW)

Compare each learner’s knowledge score with their confidence. Flag high-knowledge / low-confidence (coach) and low-knowledge / high-confidence (risk).

RUBRIC · TRANSFER (LEVEL 3)

- 0 **No mention** of use.
- 1 **Intends to** / general yes.
- 2 **One specific example.**
- 3 **Example + mentor corroboration.**

INSTRUCTION 3 · BLOCKERS (COLUMN)

Across learners who scored ≤1 on transfer, theme what blocked application; count; flag any rising module-level gap.

✓ **Knowledge and confidence are not the same signal**

The high-knowledge / low-confidence learner needs coaching, not re-training — a different fix and a different cost. Scoring them separately, on one record, is how Rosa targets the right intervention instead of repeating the whole module.

LEARNING INTELLIGENCE · STEP 3

From one reflection to the funder report.

One 120-learner cohort, worked through the four scopes.

CELL · ONE REFLECTION

```
learner L-221 · 90-day
transfer: 2 / 3
quote: “ran a structured 1:1 with my team for the first time…”
mentor: not yet corroborated
```

ROW · ONE LEARNER

L-221 — pre, post, mentor, 90-day on one ID. Knowledge +2, confidence flat, one real application example. **Net: applying it; ask mentor to confirm at day 120.**

COLUMN · TRANSFER AT 90 DAYS

```
applied on the job (n=98 reached):
yes, with example ▇▇▇▇▇ 67%
intent only      ▇▇ 21%
no               ▇ 12% · blocker: scheduling
```

The renewal answer Rosa couldn’t produce — now a single column.

GRID · KIRKPATRICK L1–L4

All four levels from one spine — reaction, learning gain, behavior at 90 days, and the business result — rendered for the funder in minutes, every figure cited.

★ **The renewal-saving number**

“67% applied the skill with a specific example by day 90; the rest named scheduling and manager support as blockers” — behavior-change evidence, cited, that completion and satisfaction could never give. That sentence is the renewal.

LEARNING INTELLIGENCE · STEP 4

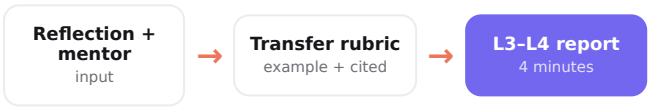
A funder answer in four minutes, not six weeks.

When pre, mid, and post share one learner ID, the report is a query — metrics and behavior-change stories together, as a live link that updates as the 90- and 180-day data arrives.

REPORTS THIS RECORD SHIPS

Pre/post deltas	by learner & segment, knowledge + confidence
Behavior change	AI-themed mentor/learner evidence, cited
Engagement flags	green/yellow/red during the cohort
Follow-up completion	90/180-day, 3× the unlinked rate
L1–L4 funder report	the full cascade, one live link

THE VISUAL THAT LANDS



Σ

Aggregate · the cascade + ROI

Reaction → learning → behavior → results on one learner ID, with program cost for a Phillips ROI layer — and reusable instruments across cohorts, so Year 2 compares to Year 1 automatically. Level 3 and 4 become default outputs.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · Completion as a proxy. A clicked-through module isn't a learned one. Measure transfer, not attendance.

2 · Asking for intent. "Will you use it?" overclaims. Ask for the instance.

3 · Designing follow-up late. Write the 90-day questions with the intake, or pre/post won't match.

CHAPTER SEVEN · PROGRAM INTELLIGENCE

Six programs, eleven funders, one standard.

The comprehensive case: a multi-program organization (or a federation of chapters and members) that has to report to every funder in the framework they agreed to. The design job is to standardize collection once, aggregate onto one record per person, and let each funder's report be a different view of the same evidence.

THE MANAGER
an M&E lead across programs

THE DECISION
report to each funder; steer the whole portfolio

THE PAIN
a different format per program, per funder

START HERE

Read **From Fragmented Data to Program Intelligence** first, then build it here — `sopact.com/ebooks/program-intelligence`

PROGRAM INTELLIGENCE · THE MINDSET SHIFT

Aisha’s six programs measure the same thing six different ways.

Aisha leads M&E for a federation: six programs across multiple chapters, each with its own logic model, its own funder, its own reporting cadence. Theory of change lives in a Google Doc, survey data in Airtable, financials as PDFs, field notes in two languages. Each funder wants a different framework. The national roll-up is a quarterly act of heroism, and 80% of her team’s time goes to reconciliation, not insight.

> *“Two chapters define ‘served’ differently, so I can’t add them up. By the time I’ve normalized everyone’s spreadsheet, the quarter’s over and I’m reporting on data that already moved on.”*
>
> AISHA · M&E LEAD, MULTI-PROGRAM FEDERATION

Where she’s stuck today

Standardization is the unsolved problem: every chapter collects differently, so nothing rolls up cleanly. And because each funder wants its own framework, the temptation is a parallel spreadsheet per funder — which multiplies the reconciliation rather than removing it.

What intelligence changes

Define one shared data dictionary and ID scheme centrally; let each chapter collect locally in its own language on the shared instruments; and bind each funder’s framework to the same connected record. The roll-up becomes a query, and each funder’s report a different narrative over one source.

→ **The reframe for this chapter**
This chapter is the others, aggregated. Case, application, grant, learning — each is one program shape; program intelligence is the **spine beneath all of them**: standardize once, report to everyone, every figure cited.

PROGRAM INTELLIGENCE · STEP 1

Standardize the dictionary; localize the collection.

The design move that unlocks everything: separate **the standard** (defined centrally, once) from **the collection** (run locally, in context). When every chapter measures against the same definitions and IDs, aggregation stops being reconciliation.

LAYER	DESIGN IT TO HOLD	THE QUESTION IT ANSWERS
CENTRAL **Data dictionary**	Shared indicators, definitions, units, and the framework each funder requires.	“Does every chapter mean the same thing by ‘served’?”
LOCAL **Multi-source collection**	Surveys, documents, interviews, offline (KoboToolbox) — in the chapter’s own language, on shared instruments.	“Can the field collect the way it works, and still roll up?”
SPINE **One ID per person**	A persistent ID minted at first contact, carrying across programs and chapters.	“Is the person in two programs one record or two?”
UP **Framework bindings**	Each funder’s framework mapped to the same indicators — not re-keyed.	“Can one record feed many funder reports?”

Each program keeps its own outcomes
Standardization isn’t one rigid model forced on everyone. Each program binds its own logic model to the same connected record — the umbrella view rolls them up without flattening what makes each program distinct.

What *not* to do
Don’t build a spreadsheet per funder. The fix is one connected record that every framework reads from — so “report to Funder A” and “report to Funder B” are two queries, not two datasets.

PROGRAM INTELLIGENCE · STEP 2

Score to the shared dictionary; output to each framework.

The instructions standardize the read and flex the output: score every program's submissions against the shared dictionary and its own logic model, flag missing data and outliers, and render per the funder's framework — all from one record.

INSTRUCTION 1 · STANDARDIZED READ (CELL/ROW)

Score each submission against the shared data dictionary + the program's logic model. Reconcile narrative to KPIs. Cite every figure; flag missing or out-of-range values.

INSTRUCTION 2 · MULTILINGUAL (CELL)

Read responses in any language; theme and score in English; preserve the original quote for the citation.

RUBRIC · REPORTING READINESS

- 0 **Missing** or non-comparable.
- 1 **Reported**, not yet reconciled.
- 2 **Reconciled** to the dictionary.
- 3 **Reconciled + cited + framework-mapped.**

INSTRUCTION 3 · RENDER (GRID)

Output the same evidence per each funder's framework (IRIS+, SDG, custom logframe). No re-keying.

✓ **One read, many renders**

The reading and reconciliation happen once, against the shared dictionary. The funder-specific framing is just a different output template over the same scored record — which is why six weeks of assembly collapses to hours.

PROGRAM INTELLIGENCE · STEP 3

TOC + reported + financial + audit = comprehensive.

This is the chapter where the four scopes become a full organizational roll-up. The Grid here isn't one report — it's the comprehensive view every funder, board, and auditor draws from.

THE FOUR SCOPES, FEDERATION-SCALE

CELL
one response, scored to the shared dictionary, cited in its original language.

ROW
one person across programs — two outcome tracks, one record.

COLUMN
one indicator across all chapters — finally comparable.

GRID
the whole federation, any funder's framework.

THE COMPREHENSIVE BUILD

TOC	outcomes vs commitments, per program
+ Reported	metrics + qualitative themes, cited, multilingual
+ Financial	cost-per-outcome from accounting
+ Audit	every figure one click from its source
= Comprehensive	**one record, any funder's framework**

WORKED · THE "SERVED" PROBLEM, SOLVED

Two chapters reported "served" differently. Bound to the shared dictionary, the AI reconciles both to the standard definition, preserves each chapter's original note for the citation, and the federation total is now **defensible** — not a guess. The same record renders SDG-aligned for one funder and IRIS+ for another, overnight.

PROGRAM INTELLIGENCE · STEP 4

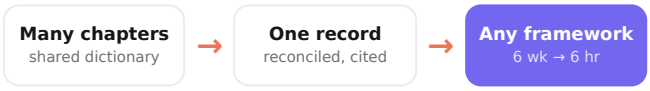
Standardize once. Report to everyone.

When collection is standardized and aggregated onto one record, every report — for every funder, board, and chapter — falls out as a query, in the language and framework each one needs.

REPORTS THIS RECORD SHIPS

Program impact	TOC-aligned, per program, multilingual
Funder report	each funder's framework, from one source
Missing-data alert	which chapters/fields are short — on the due date
Outcome variance	programs below commitment, root cause from the qual
Early-warning	dropout & performance flags as data arrives
Board / partner	executive narrative + per-chapter feedback

THE VISUAL THAT LANDS



Σ

Aggregate · the whole spine

Every prior chapter rolls up here: the case note, the application score, the grantee report, the learning transfer — one connected record per person, one shared dictionary, every figure cited. Standardize the input; the comprehensive report is a join.

THREE MISTAKES THAT QUIETLY BREAK THIS

1 · A spreadsheet per funder. Multiplies reconciliation. One record, many renders.

2 · No shared dictionary. If "served" drifts by chapter, nothing rolls up. Define it centrally.

3 · Translating twice. Collect in the local language, analyze once, render in any — don't run two translation teams.

BRING IT TOGETHER

Seven programs. One habit: ask for the evidence.

Across every chapter the craft was identical — only the program changed. Start with the decision; choose the instrument that reveals the most; ask for the example, not the opinion; write an instruction with a rubric that cites its source; and report at the scope your question needs. Do that, and a survey stops being the ceiling on what you can learn.

1

Decide first

Name what the answer will change.

2

Read the 95%

Documents & interviews, not just the survey.

3

Score with a rubric

Cited, reproducible, comparable.

4

Report at the right scope

Answer, person, question, or picture.

SOPACT · THE GUIDE

Read the eBook for your work. Then design one question the new way.

Each chapter is the hands-on next step from one guide in the library. Download the matching eBook, bring one real document or transcript, and watch it become a cited, comparable score in a walkthrough on your own data.

Download the eBooks →

sopact.com/ebooks