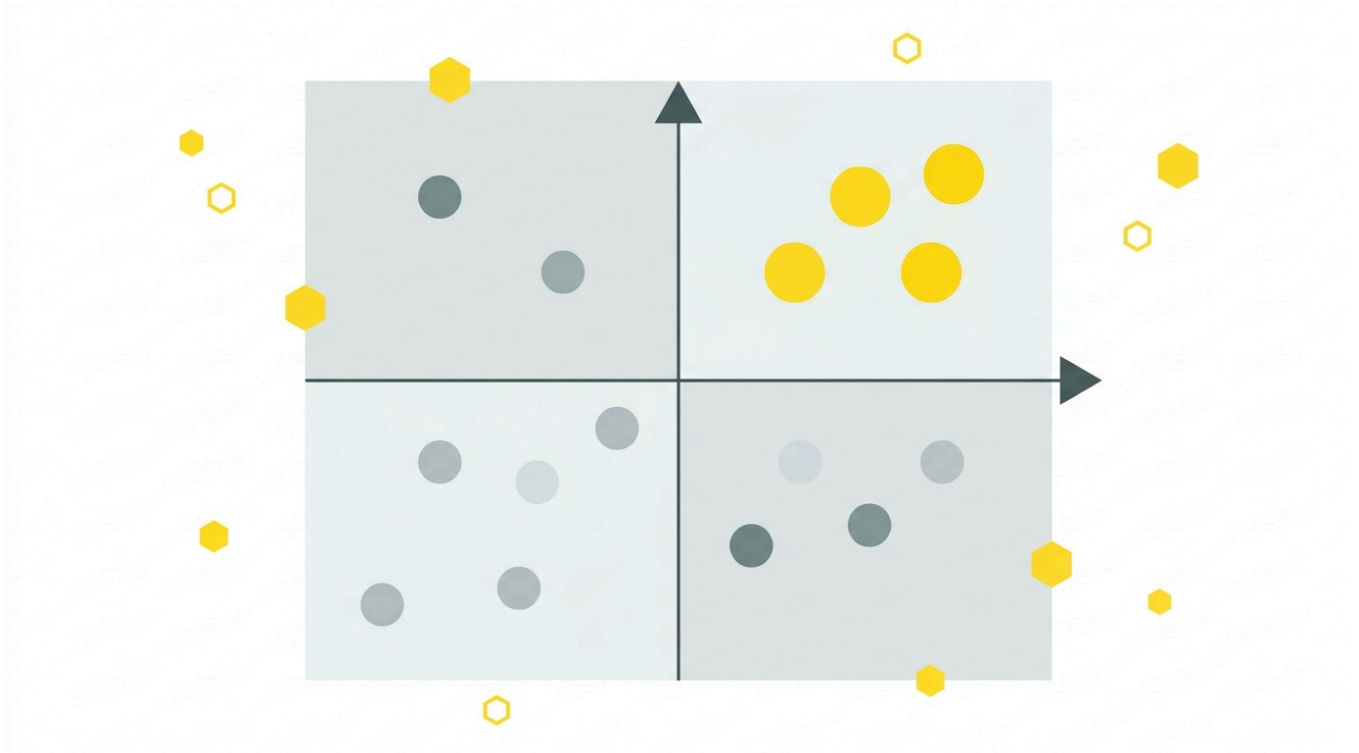




Pharma Data Platform Use Cases — Ranked by Complexity × Value

Eleven use cases the A4BEE Data Platform Blueprint engagement has seen prioritized across pharma + biotech manufacturing, quality and development orgs. Plotted on a complexity × value 2×2. Use it to sequence your platform's roadmap — quick-wins first, headline use cases last. Effort in person-months (pm) is order-of-magnitude, for a mid-size site.

Why a 2×2 instead of a feature list

Data platform programmes that fail almost always fail the same way: they pick the most ambitious use case first, run into platform foundation issues, and stall before delivering anything. The 2×2 forces you to put quick-wins (low complexity, real value) ahead of headline use cases (high complexity, high value).

The honest sequencing: **lower-left first, lower-right second, upper-left third, upper-right last.** By the time the platform survives the lower quadrants, the upper-right ones become tractable. Inverting the order is how 18-month programmes turn into 4-year programmes.

The 11 use cases at a glance

Quadrant	Use cases	When to schedule
Low complexity / Low value	Golden-batch profiling · Batch genealogy & material traceability · Environmental-monitoring trending	First 90 days — proves the pipeline + governance work
Low complexity / High value	Deviation & CAPA analytics · OEE / line-productivity benchmarking · Right-first-time batch prediction	Months 4–9 — first measurable business outcome
High complexity / Low value	Full multi-site MES/LIMS harmonization · Cross-company consortium data sharing	Months 10–18 — only if a network mandate demands
High complexity / High value	Multivariate modelling → Real-Time Release Testing · Pharmacovigilance signal detection · Model-based tech transfer	Months 18+ — the headline. Don't start here.

The quadrants

Quick-wins — low complexity, low value

Use these to prove the platform works, not to declare victory.

Golden-batch profiling

Complexity drivers: historian + LIMS data joined into reference trajectories for a known-good product. Mostly ingest + alignment; the "answers" already exist in your best historical batches.

Value drivers: baseline — a shared golden-batch overlay retires a stack of manual trend charts and gives every process engineer the same reference. Modest, but real. *Why ship it first:* it exercises every layer of the platform (ingest, contextualize, model, consume) on a problem you can validate against past data. If your platform can't do this, it can't do anything else. *Rough size:* 2–4 pm.

Batch genealogy & material traceability

Complexity drivers: link lots, materials, equipment and process steps into one queryable graph across MES/ERP/LIMS. Data-wrangling heavy, no real-time decisions, no modelling. *Value drivers:* faster investigations and recall scoping — hours instead of days when QA asks "where else did this lot go?". *Why ship it second:* it's the ideal governance and master-data workout, and the genealogy graph it builds is a dependency for almost everything upstream in the value quadrant. *Rough size:* 3–5 pm.

Environmental-monitoring trending

Complexity drivers: clean-room EM + clean-utility data trended with excursion flags. Rule-bound thresholds (Annex 1 grades), low-risk, well-bounded. *Value drivers:* retires manual Excel trend reporting, tightens excursion response, gives QA a live view instead of a monthly binder. *Why ship it third:* introduces scheduled data quality + alerting on regulated data without any decision-support risk. *Rough size:* 2–3 pm.

First real wins — low complexity, high value

These pay back the platform investment for the first time.

Deviation & CAPA analytics

Complexity drivers: cluster, trend and triage deviations and CAPAs across the site. The models are simple; the work is normalizing categorization and linking to batch/equipment context.

Value drivers: high — direct impact on QA cycle time, right-first-time, and recurring-deviation cost. A metric the plant manager reports upward every month. *Why ship it fourth:* first use case with measurable financial + compliance outcomes. By month 6, leadership sees the dashboard and signs off the next platform investment round on this case alone. *Rough size:* 3–5 pm.

OEE / line-productivity benchmarking

Complexity drivers: downtime, changeover and utilization across lines and sites. Comparison logic is straightforward; the hard part is normalizing across equipment and shifts. *Value drivers:* high — capacity and cost improvements tied to concrete numbers. Board-level metric, and a different stakeholder (Ops/COO) than the quality cases. *Why ship it fifth:* operational rather than quality, but same platform infrastructure — it broadens the platform's internal customer base. *Rough size:* 3–4 pm.

Right-first-time batch prediction

Complexity drivers: flag at-risk batches early from in-process signals. First genuinely predictive workload — but the modelling approaches (classification on historized process data) are well understood. *Value drivers:* high — scrap and rework avoided is immediate, quantifiable savings, and it's the on-ramp to the multivariate work in the headline quadrant. *Why ship it sixth:* it introduces predictive modelling on process data while the outcome (was the batch in spec?) is unambiguous and easy to validate. *Rough size:* 4–6 pm.

Slow grinds — high complexity, low value

Avoid unless mandated. These rarely pay back the engineering time on their own.

Full multi-site MES/LIMS harmonization

Complexity drivers: enormous — harmonizing schemas, master data, and workflows across sites with different systems and different validation states takes years. *Value drivers:* indirect. Real, but it accrues to the network, not to any one site's P&L, and it competes with use cases that pay back in months. *When to ship:* only when a network mandate (a global template programme, a post-merger integration) forces it. Don't volunteer for it. *Rough size:* 18+ pm.

Cross-company consortium data sharing

Complexity drivers: shared reference data / benchmarking across partners means multi-party governance, IP protection, and legal agreements that dwarf the analytics. *Value drivers:* moderate — useful for benchmarking consortia; thin for a single manufacturer acting alone. *When to ship:* when the contracts and standards exist. The data engineering is endless and the analytic payoff is late. *Rough size:* 12+ pm.

Headline use cases — high complexity, high value

The reason you built the platform. Don't start here.

Multivariate modelling → Real-Time Release Testing

Complexity drivers: MVDA + spectral (NIR/Raman) models mature enough to release product on prediction instead of end-of-line lab tests. Requires a validated platform, mature CPV, and a regulatory strategy. *Value drivers:* very high — collapses release timelines and lab load, frees working capital tied up in hold inventory. *Why ship it late:* it needs the golden-batch, genealogy and RFT foundations solid and validated first. Attempted early it produces models nobody will release against. *Rough size:* 9–15 pm.

Pharmacovigilance signal detection

Complexity drivers: NLP on adverse-event narratives + disproportionality analysis + causal inference. Confounders are everywhere; analyses that look right are often wrong. *Value drivers:* very high — patient safety, regulatory standing, and post-market surveillance. Structural and growing interest. *Why ship it late:* needs clean longitudinal data, a validated NLP pipeline, and senior data science to stay methodologically defensible. *Rough size:* 9–12 pm.

Model-based tech transfer

Complexity drivers: move validated process knowledge between sites as calibrated models, not binders — process models, design spaces, and their data lineage travelling together. *Value drivers:* very high — compresses tech-transfer timelines, the rate-limiting step for launch and second-source supply. *Why ship it latest:* it depends on the full mechanistic + data foundation and on genealogy/ontology being solid across sites. Premature attempts transfer models nobody trusts at the receiving site. *Rough size:* 9–12 pm.

A recommended 18-month sequence

Months	Quadrant	Use cases shipped
0–3	Low / Low	One quick-win — typically golden-batch profiling. Output: a working pipeline + a dashboard process engineers actually use.
4–9	Low / High	Deviation & CAPA analytics + one operational case (OEE). Output: first measurable business outcome, signs off the next platform investment round.
10–15	Foundation work	Ontology v2, genealogy graph completion, governance + NFR uplift. Not a use case — but without this, the next quadrant fails.
16–24	High / High	One headline use case — typically RTRT or model-based tech transfer. Output: the case the platform was built for.

Skip the **High / Low** quadrant unless a network mandate or regulation forces it. Most orgs that started there in years 1–2 have nothing to show for it three years in.

How this maps to the A4BEE Data Platform Blueprint

The [Data Platform Blueprint engagement](#) uses this 2×2 in the week-2 on-site workshop to align leadership on the sequencing. Stakeholders usually arrive with one of the upper-right use cases

in mind ("we want Real-Time Release Testing!"). The workshop talks them down to the lower quadrants first. Bluntly.

For the maturity context that informs whether your platform can even support the lower quadrants yet, take the [Digital Strategy Maturity assessment](#).

A4BEE · 2026