# AI Lab Transition Strategy

Complete Phase Guide — Life Sciences Wet Labs

---

This document provides the full breakdown of each phase in the AI Lab Transition framework, covering all three pillars — Technology, Organisation & Culture, and Data Supply — along with business outcomes, success metrics, and honest assessments.

**20/15 Visioneers — Strategy & IT Consultancy for Life Sciences R&D**

www.20visioneers15.com | john@20visioneers15.com

# Data Readiness

Making lab data AI-consumable. Not just digital — structured, contextualised, and queryable. This phase overlaps entirely with the Smarter Labs transition.

## TECHNOLOGY

**Instrument data capture**
Machine-readable formats — not PDFs of CSVs

**Metadata standardisation**
Consistency across assays, projects, and sites

**Experiment-result linkage**
What was done → what happened → what it means

**Ontology adoption**
BAO, ChEBI, OBI — even lightweight mapping

## ORGANISATION & CULTURE

**Leadership sponsors data as strategic priority**
Not delegated to IT. Scientists deprioritise without visible leadership commitment.

**Lab champions (2–3 per group)**
Senior scientists with peer credibility who model compliant behaviour.

**Data quality ownership explicitly defined**
Scientist, lab manager, or data steward — ambiguity guarantees poor data.

Survey: Data Practices & Attitudes — baseline before any technology deployment

■ Expect passive non-compliance: incomplete fields, free-text where structured input is required.

## DATA SUPPLY

**Audit captured vs. generated data**
Most labs lose 30–60% of instrument output — only 'final results' are recorded.

**Prospective capture design**
If Phase 1 needs environmental data for drift prediction, start capturing now.

**External ontology mapping**
Lightweight but outsized impact on future interoperability.

**Data quality baseline**
Error rates, completeness, consistency metrics — no baseline means no way to measure improvement.

## BUSINESS OUTCOMES

- Eliminates manual re-entry and transcription errors, reclaiming scientist time
- Creates a queryable institutional knowledge base — experiments findable across teams, sites, and years
- Reduces repeat experiments caused by inability to find or trust prior results
- Establishes the data asset that underpins all subsequent AI value

## SUCCESS METRICS

- >90% of instrument data in structured, machine-readable format
- Mean time from experiment completion to data availability
- % of experiments with complete metadata
- Cross-team/site data queries executed per month

- Survey delta vs. baseline on data practices & attitudes

## HONEST ASSESSMENT

This is where 80% of labs are stuck. Most 'digital transformation' projects produce data that is technically digital but practically unusable for ML or agents. If your Smarter Lab engagement hasn't solved this, the AI transition cannot start. This phase is the gatekeeper. The organisations that get through it fastest are those that have access to people who have done this before — who know which metadata fields matter, which instrument integrations are worth the effort, and which vendor promises about 'seamless data capture' will quietly fall apart at scale.

## CONTRARY VIEW

Some vendors will claim their platforms make data 'AI-ready' out of the box. This is rarely true for multi-instrument, multi-assay wet lab environments. The heterogeneity problem is severe.

# Predictive Intelligence

Deploying ML models that surface predictions, patterns, and anomalies from lab data. Humans retain all decision authority.

---

## TECHNOLOGY

**Assay performance prediction**
Yield, variability, failure probability

**Instrument drift detection**
Predictive maintenance scheduling

**Compound-property prediction**
ADMET, solubility, stability

**Experiment prioritisation / DoE**
Classical ML often outperforms deep learning on tabular lab data with small n

## ORGANISATION & CULTURE

**"Trust but verify" posture**
Leadership endorses ML predictions without mandating blind compliance.

**ML literacy training**
Scientists need to critically evaluate outputs — not build models, but ask the right questions.

**Accountability framework**
When an ML recommendation fails, who is accountable? Answer before deployment, not after.

Survey: AI Readiness & Trust — quarterly. Track confidence, autonomy concerns, ML comprehension.

■ "The model is wrong" — some cases are genuine failures, some reveal flawed human heuristics. Both must be handled constructively.

## DATA SUPPLY

**Labelled training data (bottleneck)**
Retrospective labelling of historical data + prospective capture design for new experiments.

**New data streams required**
Environmental monitoring, reagent lot tracking, operator-level variability data.

**External datasets**
ChEMBL, PubChem, BindingDB — models trained only on internal data have narrow applicability.

**Volume reality check**
Most wet lab datasets are too small for deep learning. Drive model selection from data, not aspiration.

## BUSINESS OUTCOMES

- Reduces failed experiments and wasted materials through outcome prediction
- Shifts instrument maintenance from reactive to predictive
- Accelerates compound triage — fewer compounds advanced to expensive late-stage assays
- Data-driven experiment prioritisation compresses discovery cycle times
- Quantitative visibility into operational performance for lab leadership

## SUCCESS METRICS

- Prediction accuracy vs. actual outcomes (per model, per quarter — watch for drift)
- Reduction in unplanned instrument downtime

- Scientist adoption rate — % of eligible decisions where ML was consulted
- False positive/negative rates (too many FPs kill adoption fastest)
- AI Readiness & Trust Survey quarterly trend

## HONEST ASSESSMENT

The bottleneck is not model capability — it is labelled training data. Scientists will not trust predictions they cannot interrogate. Explainability is not optional.

## CONTRARY VIEW

Many labs will get more value from simple statistical process control and dashboards than from ML models. Don't deploy ML where a control chart would suffice — it erodes trust when the complexity isn't justified. The problem is that nobody selling you ML tools will tell you this. The incentive to recommend ML when simpler methods would suffice is strong, and the only protection is independent judgement from someone whose revenue isn't tied to model complexity.

# Augmented Workflows

LLM-based agents that draft, summarise, retrieve, and recommend — but require human sign-off before any action is taken.

---

## TECHNOLOGY

**Protocol & document drafting**
SOPs, deviation reports, stability summaries from historical data

**Literature-informed experiment design**
RAG over internal + external corpus

**Natural language query over lab databases**
Cross-system data access without knowing which system holds what

**Equipment troubleshooting**
Maintenance logs + manuals via agent interaction

## ORGANISATION & CULTURE

**Explicit AI authority boundaries**
Which document types can agents draft? Documented by leadership, not negotiated ad hoc.

**AI Review Board**
Cross-functional (science, quality, IT, regulatory) — reviews use cases before deployment.

**AI output review skills training**
Agent errors are qualitatively different from human errors — confidently wrong, subtly fabricated.

Survey: Workflow Impact & Autonomy — quarterly. Include anonymous free-text for real concerns.

■ Approval fatigue: review rigour decays under time pressure. "It'll take my job": displacement concern becomes concrete and personal.

## DATA SUPPLY

**Curated RAG corpus**
SOPs, protocols, historical records — must be kept current. Stale data = stale outputs.

**Agent interaction logs**
Every prompt, output, review decision, and edit — essential for quality monitoring.

**External literature access**
PubMed, bioRxiv, regulatory guidance — version-controlled to avoid citing withdrawn documents.

**IP governance framework**
Which models see which proprietary data, where inference runs, what is retained.

## BUSINESS OUTCOMES

- Document drafting time reduced from days to minutes (starting from 80% not 0%)
- Scientists spend less time searching, more time interpreting
- Institutional knowledge accessible to new hires immediately
- Reduced dependency on "the one person who knows how this assay works"
- Faster deviation and CAPA cycle times

## SUCCESS METRICS

- Agent output acceptance rate — below 70% means creating more work than saving
- Hallucination/error rate per use case, per model

- Time-to-first-complete-draft per document type
- Review time trend (declining may signal approval fatigue — a warning)
- Scientist satisfaction scores and Workflow Impact Survey delta

## HONEST ASSESSMENT

This is where the gap between demo and production is widest. LLM agents are catastrophically unreliable when exposed to messy lab data, ambiguous SOPs, and exception-heavy workflows. Hallucination in a scientific context is a data integrity risk. Every agent output touching a regulated process must be verifiable. Closing this gap requires people who can both architect the AI system and understand the science it operates on — buying an off-the-shelf agent platform and hoping it works with your data is how most Phase 2 projects fail.

## CONTRARY VIEW

The "human in the loop" framing gives false comfort. In practice, approval fatigue sets in quickly — scientists will rubber-stamp agent outputs they don't fully review. The governance model must account for this.

# Bounded Autonomy

AI systems that make and execute decisions within pre-defined boundaries, without requiring per-decision human approval. Narrow scope only.

---

## TECHNOLOGY

**Automated parameter adjustment**
Within validated instrument ranges only

**Dynamic experiment scheduling**
Resource availability and priority optimisation

**Automated data QC triage**
Flag / hold / release based on predefined criteria

**Predictive inventory ordering**
Consumption-based reordering

## ORGANISATION & CULTURE

**Delegated decision authority**
Leadership explicitly accepts accountability for AI-made decisions.

**Role redesign**
Scientists shift to exception handler & auditor — a fundamentally different job.

**QA/Regulatory co-design**
Quality and regulatory teams architect guardrails, not just review completed systems.

Survey: Post-Autonomy Impact — per use case + quarterly. Complement with structured interviews.

■ QA/Regulatory resistance: "How do we audit a decision no human made?" Must be answered before deployment.

## DATA SUPPLY

**Real-time data feeds**
Live instrument, inventory, and scheduling integration — not batch uploads.

**High-volume decision logs**
Trigger → inputs → output → action → outcome. Auditable and retention-compliant.

**External operational feeds**
Supplier lead times, regulatory calendar, clinical trial timelines.

**Closed-loop feedback data**
Continuous self-monitoring for drift — without this, autonomous systems are unauditable.

## BUSINESS OUTCOMES

- Lab operations run closer to 24/7 — decisions happen in real time, not at morning stand-ups
- Throughput increase where decision latency is the current bottleneck
- Reduced cognitive load on senior scientists as routine-decision bottleneck
- Inventory stockouts and waste both decrease
- Governance infrastructure prerequisite for self-driving lab concepts

## SUCCESS METRICS

- Autonomous decision accuracy >95% (retrospective audit)
- Intervention rate monitored (zero = nobody is checking)
- Zero scope-creep incidents (any non-zero = governance failure)

- System uptime and fallback activation rate
- Post-Autonomy Impact Survey confidence and safety trend

## HONEST ASSESSMENT

Most wet labs will not — and should not — reach this phase for years. The labs most likely to adopt bounded autonomy first are high-throughput screening and PAT environments where the decision space is well-defined and the feedback loop is fast. For discovery biology, Phase 2 with human oversight is the realistic ceiling for the medium term. The governance frameworks required here — the guardrail design, the audit architecture, the regulatory defensibility — cannot be bolted on after deployment. They must be co-designed with the technology from the outset, by people who understand both the AI systems and the regulatory expectations they must satisfy.

## CONTRARY VIEW

Competitor pressure from AI-native biotechs may force faster adoption than traditional pharma risk frameworks would normally allow. Worth monitoring — but not worth betting a quality system on.

# Longitudinal Survey Framework

Surveys are not one-off activities. They form a measurement programme that tracks organisational readiness and surfaces resistance before it stalls deployment.

| Survey | Deployed At | Frequency | Purpose |
| --- | --- | --- | --- |
| Data Practices & Attitudes | Phase 0 baseline | Annually | Track data culture maturity and scientist burden |
| AI Readiness & Trust | Phase 1 entry | Quarterly | Monitor trust, autonomy concerns, ML literacy |
| Workflow Impact & Autonomy | Phase 2 entry | Quarterly | Detect approval fatigue, role anxiety, adoption barriers |
| Post-Autonomy Impact | Phase 3 post-deploy | Per use case + quarterly | Confidence, safety perception, role evolution |