Content overview

A short orientation before the detail begins — no science background needed

WHAT THIS IS

A worked example of how the data behind a single laboratory measurement is governed — the files, the processing rules, the settings, and the checks that together decide whether a number can be trusted, and defended, years after it was produced.

WHY I BUILT IT

To show how I work rather than describe it. The subject is scientific. The job is not. The job is deciding what a record must contain, which system owns it, what is allowed to change it, and how anyone would notice if something quietly went wrong.

WHAT IS INSIDE

- The pipeline that turns a raw file into a reported number, and why every step is fixed in advance.
- The four things that can change the answer, and how each one is locked and tracked.
- The dictionary saying what every record must contain, who owns it, and what may change it.
- The test: could anyone reconstruct a released result, years later, from what was kept?

HOW TO READ IT

Technical terms are explained where they appear, and every figure is either real or marked as representative. Nothing here names an employer, a product, an instrument, or a software supplier.

Kalif Shear · khshear@outlook.com · linkedin.com/in/kalifshear

Analytical Data Pipeline and Rule Governance

Canonical files, deterministic processing, locked configuration, drift monitoring, multi-site harmonisation

WORK SAMPLE · EXHIBIT 1 OF 6

Illustrative. No instrument, software vendor, employer, or product named.

DETERMINISTIC PROCESSING PIPELINE

a versioned rule set, applied identically to every file — not a sequence of human decisions

RAW ACQUISITION FILE → ACQUISITION SYNC GATE → SINGLE-EVENT GATE → VIABILITY GATE → TARGET POPULATION GATE → REPORTED RESULT

The order is fixed and the thresholds are versioned. Two analysts running the same file get the same number — which is the only condition under which a trend means anything.

GOVERNED DATA OBJECTS

each one versioned, owned, and audit-trailed — changing any of them changes the result

CANONICAL RAW FILE

The acquisition file is the record of origin and is never edited. Its embedded metadata carries the acquisition parameters, detector settings, and reagent lot identifiers that make the run reproducible.

COMPENSATION MATRIX

Derived from defined single-stain and minus-one controls, then locked to the method version. Adjusting it after acquisition without an audit trail is a 21 CFR Part 11 finding, not a preference.

GATING RULE SET

The pipeline above, as a named and versioned package. Superseding it forces a change-impact assessment and, where the result definition moves, requalification.

ACCEPTANCE CRITERIA

The thresholds a result is judged against, linked to exactly one active specification record and carried with the method version rather than with the analyst.

MONITORING THE DEPENDENCIES

the method rests on hardware and on other sites, and both drift

INSTRUMENT PERFORMANCE DRIFT

Daily calibration output is written to the data system and trended on control charts. Signal intensity and coefficient of variation are tracked per detector, so a degrading optical or fluidic path is flagged as a trend before patient material is committed to an instrument that has quietly moved.

Rule: an instrument outside its control limits cannot be selected for a release run.

MULTI-SITE HARMONISATION

Raw files from every site land in one repository through an automated ingestion pipeline, with tiered retention by lifecycle stage. A single global rule set is applied centrally rather than locally, which is what makes a result from one site comparable to a result from another instead of merely similar.

Rule: sites contribute data and context. They do not each own a private copy of the rules.

WHAT THIS ARCHITECTURE PREVENTS

each of these is a real failure mode, and each is prevented by a rule rather than by care

Two analysts, two numbers
Fixed gate order and versioned thresholds remove the judgement call.

A silent configuration edit
Locked, audit-trailed config objects make a quiet change impossible.

Drift found by accumulation
Daily control charts flag a degrading dependency before material is committed.

Sites that agree by coincidence
One central rule set, applied to every site's raw files.

THE SAME ARCHITECTURE IN A CONTENT CATALOG

Canonical raw file, never edited
Source asset and its embedded technical metadata

Versioned deterministic rule set
Processing and derivation rules applied uniformly, not per team

Locked configuration, audit-trailed
Config that changes derived values, versioned rather than tweaked

Central rules, distributed contribution
One global definition, many sites and systems feeding it

Kalif Shear · khshear@outlook.com · linkedin.com/in/kalifshear

Data Dictionary and Rule Register

What must be present on every record, what governs it, and how a change to the rules is tracked

A. CANONICAL RAW FILE

required embedded metadata — a file missing any of these cannot be processed

FIELD	WHAT IT CARRIES	SOURCE OF TRUTH	MUTABILITY
`run_id`	Unique acquisition event identifier	Instrument	**Immutable**
`method_id + version`	The exact method version the run was executed under	Method master record	**Immutable**
`acquisition_params`	Detector settings and acquisition configuration at run time	Instrument	**Immutable**
`reagent_lot_refs`	Lot identifiers for every critical reagent used	Inventory system	**Immutable**
`operator_id`	Qualified analyst who executed the run	Training system	**Immutable**
`sample_ref`	Patient or product identifier and its chain-of-custody context	Chain-of-custody record	**Immutable**
`instrument_id`	Physical unit, tied to its calibration history	Asset register	**Immutable**

B. RULE SET VERSION REGISTER

the processing rules are an entity too, with their own versions and impact class

RULE SET	CHANGE	IMPACT CLASS	REQUALIFICATION
`GATE-v1.0`	Initial transferred rule set	Baseline	Complete
`GATE-v1.1`	Threshold clarified; no change to result definition	Minor	**No**
`GATE-v2.0`	Viability gate criteria revised after investigation	**Major**	**Required**
`COMP-v1.0`	Compensation matrix locked to method version	Baseline	Complete
`SPEC-link`	Active specification record confirmed unchanged	None	**No**

C. WHAT IS TRENDED, AND WHAT TRIPS AN ACTION

monitoring is defined per object, not left to whoever is looking at the charts

METRIC	TRACKED PER	LIMIT BASIS	ACTION ON BREACH
`signal_intensity`	Detector, per instrument, per day	Historical control limits	Instrument blocked
`precision_cv`	Detector, per instrument, per day	Historical control limits	Instrument blocked
`nonreportable_rate`	Method version, per period	Method acceptance criteria	Case opened
`cross_site_delta`	Site pair, per method version	Pre-agreed equivalence margin	Harmonisation review
`metadata_completeness`	Run, at ingestion	All required fields present	Run rejected at ingest

WHY A DICTIONARY AND A REGISTER, RATHER THAN A DASHBOARD

A dashboard reports that records are incomplete. A dictionary makes an incomplete record impossible to create, and a register makes a rule change impossible to make quietly. The first is observability; the second and third are governance. Everything above is enforced at write time and at state transition — which is why the health numbers in Exhibit 4 can be trusted at all.

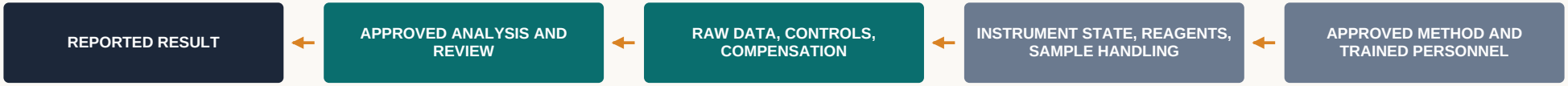
Kalif Shear · khshear@outlook.com · linkedin.com/in/kalifshear

Reconstructing a Released Result

The reconstructability test — what must be retrievable, and what the owner must be able to answer with evidence

THE LINEAGE CHAIN, READ BACKWARD

every reported value depends on a chain, and the chain is only as good as its weakest recoverable link



If any link is missing, uncontrolled, or unrecoverable, the result is not defensible — whether or not it passed. Passing and defensible are different properties.

RETENTION MANIFEST

what must be retrievable for any single released result, years later

01	Original raw list-mode files, including controls, compensation controls, and failed runs
02	Sample identifier with its chain-of-identity and chain-of-custody context
03	Instrument ID, configuration, calibration and daily QC results, acquisition timestamp
04	Acquisition template version and the settings actually used, not those intended
05	Reagent panel identity, lot, expiry, preparation records, and qualification data
06	Compensation data and the source controls it was derived from
07	Analysis template version, gate hierarchy, gate statistics, derived calculations
08	Any manual adjustment with justification, original analysis kept beside the reviewed one
09	Analyst, reviewer, approver, timestamps, signatures, and a reviewable audit trail
10	Deviations, investigations, and the final reported result as issued

WHAT THE METHOD OWNER MUST ANSWER, WITH EVIDENCE, ON DEMAND

ownership is of the state of control, not of a folder of historical documents

What does this measure, and why is it fit for that decision?
Target profile, intended use, acceptance criteria, known limitations, and the basis for every gate and control.

What is the approved state today?
Current procedure, panel, reagent specifications, instrument configuration, analysis template, and software versions.

How has it changed?
Version history linking each change to its control record, risk assessment, comparability data, approval, and training impact.

Is it still performing as expected?
Trending of controls, instrument QC, invalid and repeat rates, precision, and analyst, instrument, and site effects.

Can every released result be reconstructed?
The manifest above, retrievable end to end for any sample or lot on request.

Does a proposed change need bridging?
Assessed on whether the change could move the measurement, not on whether the procedure wording changed.

WHERE THIS EXPECTATION COMES FROM

ICH Q14 treats an analytical procedure as a living system whose knowledge is managed across the product lifecycle. ICH Q10 defines that knowledge management explicitly. Compendial flow-cytometry guidance treats sample handling, staining, instrument setup, collection, analysis, storage, and QC as parts of the method rather than administration around it, and expects a stored result to remain traceable to its original file, settings, and QC. FDA cell-therapy guidance expects the panel and the gating strategy defining each population to be described. Data integrity is judged on ALCOA+.

Kalif Shear · khshear@outlook.com · linkedin.com/in/kalifshear