

Science *Architecture*

How the Exoplanet Codex checks measurements, promotes results, freezes references, and preserves the scientific record.

This document explains how the Exoplanet Codex decides which scientific results it can trust. Pipeline details appear only when they explain a scientific safeguard. Star-specific abundance values belong in that star's Science Product Package, not here.

The two-engine floor

A result from one measurement engine can be internally consistent and still be wrong. The Codex therefore runs two independent engines for every element that has a validated model atom. Their agreement is an honesty check that neither engine can provide alone.

Engine A is the 1D-NLTE production path. It measures an equivalent width or uses synthesis-v2, then applies a per-line NLTE correction from a one-dimensional correction table.

Engine B performs NLTE synthesis inside Turbospectrum with the Gerber 2023 deck. It uses native departure coefficients for each line where a validated atom exists.

`pipeline/engine_selection.py:select_element` compares the engines line by line.

When both provide acceptable measurements, the element result uses inverse-variance weighting. Cross-engine spread remains a separate diagnostic. `CROSS_ENGINE_MIX_GATE`

raises when per-line winners produce a material conflict at the element level. A missing NLTE grid stops the pipeline; it never triggers a silent LTE fallback.

SECTION 02

The three-gate promotion

An element does not become `gold` because one comparison looks good. Promotion requires three independent checks.

- Gate 1 tests whether Engine B recovers the external INSPECT or Amarsi anchor within 0.05 dex.
- Gate 2 applies the validate-don't-tune firewall to the model atom.
- Gate 3 checks whether the absolute cross-engine difference, $|dCE|$, lies within the ratified band for the two-engine artifact.

The rule is strict. All three gates must pass. A missing gate is a failure, not an abstention.

SECTION 03

Ratified freeze tiers

Every element has one of three freeze tiers.

- `gold` means the value is approved and frozen in the active reference.
- `owed` means work or approval remains before the value can be frozen.
- `data-gap` means no value has been measured.

`owed-HELD` means the verdict contains a measured value that has not been promoted.

`owed-BLANK` means no measured value exists. Ryan ratifies each element's tier. The pipeline enforces that decision; it does not make it.

SECTION 04

The blank-cause honesty tripwire

A historical failure mode allowed a verdict row to say that no independent-gf line survived even when the tracker recorded surviving lines and a measured value. The text made a real measurement look blank.

`pipeline/provenance_honesty.py:_assert_blank_cause_is_honest` rejects that contradiction. The phase_c verdict emitter and gold-reference builder call the same guard. This makes the bad state unrepresentable before it can enter a frozen artifact.

SECTION 05

The disposition report

The disposition report is the final straggler sweep before a freeze. It reads the phase_c verdict, gold reference, two-engine artifact, and element status tracker. It assigns each element one disposition: PASS held, PASS candidate, provisional flip, owed with cause, or data-gap.

The report is machine-generated at `data/audit/element_disposition_report.json`. Analysts do not edit it. The generator runs before every freeze so unresolved elements are visible before the reference becomes immutable.

SECTION 06

The generated element status tracker

The tracker was once edited by hand. That allowed the table to drift away from the evidence it summarized.

It is now generated from the phase_c verdict, `config/physics_regime_rya400.yaml`, and `data/audit/element_status_tracker_editorial.yaml`. An explicit `verdict` column keeps the scientific verdict separate from the freeze tier. A `--check` run compares the committed file with a fresh generation. Any hand edit becomes a build failure.

SECTION 07

The four-noun naming convention

A state artifact's filename should reveal how it may change.

- A Catalog is a master enumeration.
- A Register holds mutable state facts.
- A Tracker records work or progress.
- A Reference holds frozen, validated truth.

`SEQUENCE.md` is a narrative overlay, not a fifth state noun. “Gold” is reserved for a Reference modifier.

SECTION 08

The read-set discipline

Every working session begins with `LEDGERS.md`. That startup index points to the register, tracker, catalogs, holdings manifest, and `SEQUENCE.md`.

The register-freshness guard blocks a pull request that changes a state surface without updating `CODEX_STATE_REGISTER.md`. The ledger-consistency guard compares the tracker, verdict, and gold reference. It reports undocumented disagreements in continuous integration and becomes blocking at merge.

SECTION 09

Sirius-only compute and continuous integration

Sirius, an HP ProBook 450 running the project's reference software stack, is the authoritative production runner. Abundance computations run there. The Mac is for research, development, and pull-request review.

Continuous integration also runs on Sirius. Twenty-four test files require iSpec. Ephemeral runners therefore fail 22 percent of the suite. `CI/test` is required before a pull request can merge to `main`. The repository uses merge commits so the landing record remains explicit.

The validate-don't-tune firewall

A model atom or correction grid must never be tuned to reproduce a trusted anchor. The physics must produce the anchor. The anchor is the test, not the target.

The corroboration-accept standing rule provides a bounded alternative when ionization balance stops on precision. Model vintage also matters. Titanium and manganese moved to ab-initio atoms after the audit. Iron, calcium, chromium, barium, and strontium remain on scaled-Drawin atoms with explicit annotations because the current effect is benign or not material.

Corrections carry idempotency guards. A correction that has already been applied must be recognized as complete. Re-running the pipeline cannot apply it twice.

DOCUMENT BOUNDARIES

The Science Architecture describes framework-level decisions. The Method page describes the general measurement pipeline. The Glossary defines terms. A star's Science Product Package carries its values, diagnostics, and caveats.