

► INDEPENDENT METHODOLOGY · STRIX · VERSION 1.2

The **evidence** layer behind Kestrel.

The data-extraction engine that turns mining disclosures into a structured, audit-traceable dataset.

VERSION

1.2

COMPANION

Kestrel Methodology 4.0

AUDIENCE

Analysts, researchers, and reviewers
using Kestrel

STATUS

Living document — describes
methodology, not a data snapshot

Contents

01	Executive summary	>
02	The problem Strix solves	>
03	Design principles	>
04	What Strix produces: four data layers	>
05	The canonical metric registry	>
06	How the data is produced	>
07	Validation: where principles become invariants	>
08	Provenance and auditability	>
09	Governance and the limits of automation	>
10	Coverage	>
11	Limitations and honest disclosures	>
12	How an analyst should read Strix data	>

APPENDICES

A	Observation data dictionary	>
B	Domains and dimensions	>
C	Glossary	>

Executive summary

Strix is the data-extraction and standardisation engine behind **Kestrel**. It turns the sustainability disclosures that mining companies publish — ESG factbooks, sustainability reports, and annual reports — into a structured, audit-traceable dataset that an analyst can interrogate with confidence.

Strix exists to support one hard thing: **honest cross-company comparison**. Two companies can both report "Scope 1 emissions" and mean materially different things — different organisational boundaries, different global-warming-potential sets, different assurance levels, different treatment of joint ventures. Most data products quietly paper over those differences to produce a clean-looking league table. Strix does the opposite: it captures the number *and* the conditions under which the number was produced, and it refuses to adjust one company's figure to match another's.

The design rests on four commitments that are enforced in code, not just asserted in prose:

1. **Numbers are never silently adjusted.** Every value is stored exactly as the company reported it. Unit conversion is *additive* — the original value and unit are always retained alongside the converted one.
2. **Every observation is traceable to its source.** Each data point carries the source document and the precise location within it — a spreadsheet cell, or a document page — so any figure can be checked against the original disclosure.
3. **"Unknown" is a real answer.** Where a company is silent on something (e.g. which boundary basis it used), Strix records *unknown* rather than guessing. Inference is never disguised as fact.
4. **Comparability is surfaced, never assumed.** The methodological differences, restatements, and quietly dropped commitments that make naïve comparison misleading are first-class data, not footnotes.

Strix covers the major listed mining companies — diversified majors and gold and base-metal producers — across many years of disclosure, organised into four data layers (numeric observations, methodology, targets, and operational events) spanning every major sustainability domain. This document explains how that dataset is produced and how it should be read. It is deliberately a *methodology* paper, not an implementation manual: it describes the data model, the guarantees, and the controls, but not the internal extraction recipes.

02 — SECTION

The problem Strix solves

Sustainability data looks comparable and usually isn't. The failure modes are well known to anyone who has tried to build a peer comparison by hand:

- **Boundary mismatch.** Company A reports emissions on an *operational-control* basis (100% of assets it operates). Company B reports on an *equity-share* basis (its ownership percentage of each asset). The headline numbers are not on the same footing, and nothing in a spreadsheet cell tells you so.
- **Methodology drift.** A company changes how it calculates a figure — a new emission factor, an inflation correction, a boundary change — and restates several prior years. The new report shows a clean trend; the break is invisible unless you hold last year's report next to this year's.
- **Quietly dropped targets.** A 2030 commitment that appears in three consecutive reports simply stops being mentioned. No retraction, no announcement — it just disappears from the disclosure.
- **Unit and label ambiguity.** "Water" can mean withdrawal, consumption, or discharge; freshwater or total; thousands of cubic metres or megalitres. The same metric name hides several different measurements.

- **Provenance loss.** By the time a number reaches a comparison chart, the trail back to the page or cell it came from is usually gone — so it can't be verified, and errors can't be caught.

A naïve pipeline optimises these problems away by normalising aggressively and discarding context. That produces a tidy dataset that is quietly wrong. Strix treats the context as the product.

03 — SECTION

Design principles

These principles are the credibility model. They are encoded in the data schemas and enforced by the validation layer, so they hold across every company and every ingest — they are not best-effort guidelines.

3.1 Never silently retrofit values

When two companies report on different bases, Strix does **not** convert one to match the other. Both are stored on their stated basis, each tagged with that basis, and the difference is exposed to the analyst. Adjustment, where an analyst wants it, happens transparently in the view layer — never silently in the data.

3.2 Preserve the original, always

Canonical-unit conversion is **additive**. The observation schema makes the original value and unit *required* fields, and the converted values are stored as *separate* fields. Every observation therefore keeps:

- `value_original` and `unit_original` — exactly as disclosed, never modified; and
- `value_normalised`, `unit_canonical`, and `conversion_factor` — the converted value, the canonical unit, and the exact factor applied.

Because the original is mandatory and conversion only ever adds fields, the disclosed figure can never be lost or silently overwritten, and every conversion is reversible and auditable.

3.3 "Unknown" is first-class

Schemas permit a null/unknown value for every factual field, and the extraction step is instructed to return *unknown* when the source does not state something — never to infer. The boundary `basis` field, for example, carries an explicit `unknown` value, and where Strix applies a documented default assumption it sets a separate `basis_inferred` flag so reviewers can see exactly where a value was assumed rather than read. *Unknown* is used on real figures whenever a company is genuinely silent — it is a recorded answer, not a theoretical option.

3.4 Provenance to the source cell or paragraph

Provenance is mandatory in the schema. Every observation is required to carry its source document plus the precise location within it: for spreadsheet sources, the sheet and the **exact cell containing the value**; for document sources, the page. The verbatim source text the model relied on is retained in the extraction record for audit. A reviewer can take any figure and go straight to the cell or paragraph it was read from.

3.5 One canonical registry is the single source of truth

Strix never invents metrics on the fly. Every figure is mapped to a stable metric definition in a controlled registry (see §5). A figure that cannot be mapped to a known metric is routed to a review queue — it never silently enters the dataset under an ad-hoc name.

3.6 Restatements and dropped commitments are kept, not overwritten

When a company restates a prior-year figure, Strix does not overwrite the old value. The restatement is recorded as its own fact — original value, restated value, the documents both came from, and how the change was detected — so the break in the series is visible rather than smoothed away.

What Strix produces: four data layers

Strix models disclosure as four distinct layers, each with its own schema and its own provenance. Keeping them separate is what makes honest comparison possible — a number means little without the conditions attached to it.

01 Observations The numeric facts — one row per company × metric × year × dimension × basis.	02 Methodology How the number was produced: boundary, GWP set, assurance, base year.
03 Targets Forward-looking commitments, with full revision history and dropped-target detection.	04 Operational events The corporate changes that explain year-over-year movement.

A fifth layer — comparability — is derived at query time (§4.5), never stored as fixed columns.

4.1 Observations — the numeric facts

One row per **company × metric × year × dimension × basis**. This is the quantitative backbone. Each observation carries:

- the original value and unit, and the canonical conversion alongside;
- the boundary **basis** (operational control / financial control / equity share / unknown);
- its dimensional breakdown (group total, by business unit, by country, by asset, by source type, by category);
- full source provenance; and
- an extraction-confidence score and any caveats the extractor flagged.

Observations span every major sustainability domain — among them GHG emissions, energy, water, safety, workforce, waste, land and rehabilitation, air emissions, community, tax, occupational health, and governance.

4.2 Methodology — how the number was produced

One record per **company × year × metric family**. This is the context that makes the numbers comparable (or that reveals they aren't): boundary basis, global-warming-potential set, assurance level, base year, restatement policy, and the treatment of joint ventures and contractors. Because methodology is interpretive rather than purely numeric, it is treated as a contestable layer (see §9).

4.3 Targets — forward-looking commitments

Each target is captured with its full revision history, so an analyst can see not just the current commitment but how it has changed over time. Critically, Strix detects targets that *disappear*: by diffing each year's disclosed commitments against the prior year, it surfaces **quietly dropped targets**, each with the year it was last seen, the year it vanished, and the verbatim commitment text.

4.4 Operational events — what explains the movement

Corporate changes — acquisitions, divestments, closures, restarts, boundary changes — that explain year-over-year shifts in the underlying numbers. Without this layer, a 20% drop in emissions caused by a divestment looks identical to one caused by genuine decarbonisation.

4.5 The fifth, derived layer — comparability

Comparability flags, progress-against-target calculations, dropped-target detection, and restatement detection are **always derived at query time, never stored as fixed columns**. This is a deliberate architectural choice: derived judgements stay consistent with the underlying facts and can be recomputed as methodology understanding improves, rather than going stale in a column. Restatements, for instance, are surfaced this way — each linking the original and restated value, the documents involved, and the basis for the detection.

The canonical metric registry

The registry is the single source of truth for *what Strix tracks and how*. It is the anchor that makes a "Scope 1 emissions" figure from one company line up with the same concept from another.

Each metric definition specifies:

- a **stable identifier** used as the foreign key everywhere downstream;
- the **canonical unit** and the accepted unit aliases that map to it;
- the **valid bases** the metric may legitimately be reported on;
- the **valid dimensions** it may be broken down by;
- **standard cross-references** (GRI / SASB / ESRS / ICMM), so a metric can be tied back to the disclosure frameworks it derives from;
- the **methodology fields most critical to comparability** for that metric; and
- **domain caveats** — the specific reasons cross-company comparison of this metric can mislead.

Metric names that appear in source documents are *always* mapped to a registry identifier; anything unmapped goes to a review queue rather than into the dataset. This is what guarantees that a query for a metric returns the same concept across every company, regardless of what each company called it. The registry expands as new metrics and domains are added; it is the controlled vocabulary the whole pipeline is anchored on.

Strix treats the **context**
as the product.

How the data is produced

Strix runs a staged pipeline, deterministic wherever determinism is possible and using large language models only for the genuinely linguistic and structural-judgement steps. The stages below describe *what* each step guarantees; the internal recipes (prompts, model configuration, and document-parsing heuristics) are intentionally out of scope for this document.

STEP 01 Source normalisation Every sheet and page reshaped to one canonical form, each value tagged with its cell.	STEP 02 Routing Data tables separated from prose; each routed to the right extractor.	STEP 03 Extraction A frontier model fills the schema under a strict structured-output contract.	STEP 04 Validation A deterministic gate enforces the invariants before anything publishes.
--	---	---	--

Source normalisation. Spreadsheets and documents are first converted to a single canonical internal shape. Spreadsheets — which arrive in wildly different native layouts across companies — are normalised so that every data sheet is presented the same way, with each numeric value annotated with the exact cell it came from. Document sources are converted with page tracking preserved. The result is that downstream extraction sees a consistent structure and never has to compute a cell reference itself — it copies the provenance through verbatim.

Routing. A lightweight classification step identifies which parts of each source are likely to contain which metrics, and separates genuine data tables from prose-and-definitions content (which is routed to methodology extraction, not numeric extraction). This keeps the expensive extraction step focused and prevents definitional text from being mis-read as data.

Extraction. The relevant slice of each source, together with the relevant slice of the registry, is passed to a frontier language model under a strict **structured-output contract** — the JSON schema is enforced at the tool-call layer, not merely described in a prompt. The model returns figures conforming to the appropriate schema, copying provenance through and reporting its own confidence. Low-confidence or schema-failing extractions are retried once under a stronger configuration before being routed to review.

Validation. Every extracted record passes through a deterministic validation gate (see §7) before it is allowed near the published dataset. Validation is where the design principles stop being aspirations and become enforced invariants.

The pipeline is built so that one ingest run corresponds to one source document covering whatever fiscal years that document discloses — including multi-year factbooks that publish a decade of data in a single workbook.

07 — SECTION

Validation: where principles become invariants

Before any record is published, it passes a deterministic validation gate. The checks that run in code include:

1. **Schema conformance.** The record must satisfy its JSON schema exactly — required fields present, enumerations respected, types correct. Anything that fails is rejected, not coerced.
2. **Registry mapping.** The `company_id` must exist in the controlled company vocabulary and the `metric_id` must exist in the metric registry. An unknown metric cannot enter the dataset; it is routed to review.
3. **Unit canonicalisation.** Every reported unit must resolve to the metric's canonical unit via the registry. The conversion factor is computed and stored alongside the original. A unit that doesn't resolve is rejected rather than guessed. Ratio/intensity metrics, which have no single canonical unit, are handled explicitly rather than forced through conversion.

4. **Anomaly detection on the source data.** Structural sanity checks flag figures that look irregular *in the source itself* — implausible magnitudes, negative values where impossible, year values outside a sane range, apparent silent unit shifts. Crucially, this flag describes **the disclosure**, not a Strix error: the number is still captured faithfully and routed to a human to determine whether the source is wrong or reality genuinely shifted.

Anything that fails validation is sent to a review queue and logged — **never silently dropped**. Failures are inputs to improving the registry and prompts, not data that quietly disappears.

08 — SECTION

Provenance and auditability

Provenance is the feature. Every other guarantee depends on an analyst being able to get from a number in Kestrel back to the exact place it was disclosed.

- **Spreadsheet sources** carry `source_doc + source_sheet + source_cell`, where the cell is the one holding the numeric value itself — not the row label, not the metric name, not the unit. This convention is strict precisely because it is what a reviewer verifies against on audit.
- **Document sources** carry `source_doc + source_page`, with the verbatim quote the value was read from retained in the extraction record.

The practical consequence for an analyst: no figure in Kestrel is a black box. If a number looks surprising, the trail to verify it is one lookup away. Because structured factbooks yield the cleanest, most cell-traceable data, they are the backbone of the numeric layer; document-sourced figures carry page-level provenance.

Governance and the limits of automation

Strix is candid about which of its outputs are mechanical and which are interpretive, and it treats them differently.

Every record carries a review state — `auto`, `needs_review`, `reviewed`, or `rejected` — and an explicit confidence score. The governing principle:

- **Observations** are numeric and structurally checkable. A high-confidence observation that passes every validation check can publish automatically; a lower-confidence one is flagged for review.
- **Methodology, targets, and operational events** are interpretive — they involve reading intent and nuance from prose. These are **contestable layers**: they are designed to reach an analyst with their confidence and provenance attached, for scrutiny rather than as settled fact.

This is a deliberate boundary. Strix automates the part that is safe to automate — faithful, traceable transcription of numbers — and keeps a human in the loop for the part that requires judgement.

On evaluation, in the interest of candour: Strix maintains a hand-labelled golden corpus of company-years used as ground truth, and the project sets explicit acceptance bars for extraction quality (high precision on observations, high field-level accuracy on methodology, high recall on targets). Validating a change today means re-running extraction against that corpus and diffing the output against the ground truth. We deliberately do not quote headline accuracy percentages, because a continuous, automated regression measurement is part of the roadmap rather than something we run on every change today — and we would rather under-claim than report a number we cannot stand behind. The deterministic controls described in §7 and the review gates above are what we rely on in the meantime, and they are real and enforced.

Coverage

Strix covers the world's major listed mining companies — diversified majors and gold and base-metal producers, including names such as BHP, Rio Tinto, Glencore, Vale, Anglo American, Newmont, and Barrick — across many years of disclosure and every major sustainability domain. The four data layers (observations, methodology, targets, events) are populated for each company to the extent the company discloses.

Rather than reproduce a coverage table that would date the moment another company is ingested, **current coverage is shown live in Kestrel** — the exact roster of companies, the number of observations, and the reporting-year range update automatically with every ingest. This document describes the methodology those figures are produced by; the figures themselves belong on the live platform.

A few structural facts that do not change with coverage:

- Data is modelled in **four layers** plus a fifth, derived comparability layer (§4).
- Observations support **six dimensional breakdowns**: group, business unit, country, asset, source type, and category.
- Every metric is anchored on the **canonical registry** and cross-referenced to GRI, SASB, ESRS, and ICMM.

Limitations and honest disclosures

Trust is built as much by what a data product refuses to claim as by what it delivers.

- **Strix reports what companies disclosed; it does not audit it.** All inputs are publicly disclosed figures. Strix faithfully captures and structures them — including, where flagged, figures that look irregular in the source — but it does not independently verify that a company's reported number is correct. Where companies obtained third-party assurance, that fact is captured in the methodology layer; Strix itself is not an assurance provider.
- **Coverage is uneven by design, and visibly so.** Disclosure depth varies enormously between companies and across years. Thinly disclosed domains are thinly populated — that reflects the state of disclosure, not a processing gap — and earlier years are sparser than recent ones. The data shows you where the disclosure is thin rather than papering over it.
- **The numeric layer is predominantly factbook-sourced.** Structured ESG factbooks yield the cleanest, most cell-traceable data, so they dominate the observation layer; narrative-PDF figures are a smaller share. Methodology, targets, and events draw more heavily on narrative sources.
- **Interpretive layers carry uncertainty.** Methodology, targets, and events involve reading meaning from prose. They are published with confidence scores and provenance precisely so that uncertainty is visible, and they are treated as contestable (§9).
- **Comparability is a tool, not a verdict.** Strix surfaces the differences that make a comparison fair or unfair; it does not deliver a single "comparable / not comparable" stamp. The analyst makes the call, with the differences laid out.

- **The registry is still expanding.** New metrics and domains are added over time; absence of a metric from the dataset may mean it is not yet in the registry, not that no company discloses it.
-

12 — SECTION

How an analyst should read Strix data

A short practical guide to using the dataset well:

1. **Read the basis before comparing.** When two companies' figures differ, check the `basis` first. An operational-control figure and an equity-share figure are not directly comparable, and the field tells you which you're looking at.
2. **Check for restatements before trusting a trend.** A smooth multi-year trend may conceal a methodology change. Strix's restatement layer shows you where a company revised history.
3. **Read events alongside year-over-year movement.** Before attributing a jump or drop to performance, check whether an acquisition, divestment, or boundary change (the events layer) explains it.
4. **Treat dropped targets as a signal.** A commitment that disappears from disclosure is often more informative than one that is restated.
5. **Use provenance when a number surprises you.** Every observation links to its source cell or page. If a figure looks wrong, verify it at source before drawing a conclusion.
6. **Respect "unknown."** An `unknown` basis or a null field is information — it means the company did not disclose it. Don't fill the gap with an assumption the company itself didn't make.

Observation data dictionary

Selected fields.

FIELD	MEANING
company_id	Controlled identifier; must exist in the company registry.
metric_id	Controlled metric identifier; must exist in the metric registry.
report_year	The reporting period the figure pertains to (not the publication year).
value_original / unit_original	The value and unit exactly as disclosed. Never modified.
value_normalised / unit_canonical / conversion_factor	The canonical conversion and the exact factor applied — stored <i>alongside</i> the original.
basis	Boundary basis: operational control / financial control / equity share / unknown.
basis_inferred	True when basis was set from a documented default rather than an explicit source statement.
entity_dimension / entity_dimension_type	The breakdown this row represents (e.g. a specific country, business unit, or asset).
source_doc / source_sheet / source_cell / source_page	Provenance: the document and the exact location of the value within it.
restated / restated_from_value / restated_from_doc	Restatement linkage to the previously published value.
extraction_confidence	Model-reported confidence in the value (0–1).
data_reporting_anomaly / ..._note	Flags that the <i>source disclosure</i> looks irregular — not a Strix error.
state	Review state: auto / needs_review / reviewed / rejected.

B — APPENDIX

Domains and dimensions

Sustainability domains anchored in the registry include GHG emissions, energy, water, safety, workforce, waste, land and rehabilitation, air emissions, community, tax, occupational health, governance, and biodiversity. The registry is the authority on the current set and continues to expand.

Dimensional breakdowns (6): group total, business unit, country, asset, source type, category.

C — APPENDIX

Glossary

Basis (boundary basis)

The organisational boundary a figure is reported under. *Operational control*: 100% of assets the company operates. *Equity share*: the company's ownership percentage of each asset. *Financial control*: assets the company controls financially. These are not interchangeable.

Restatement

A company revising a previously published figure for a past year. Strix keeps both values.

Dropped target

A forward-looking commitment that appears in prior disclosures and then silently disappears.

Operational event

A corporate change (acquisition, divestment, closure, restart, boundary change) that explains movement in the underlying numbers.

Canonical unit

The single unit Strix converts each metric to for comparison, while always retaining the original.

Provenance

The traceable link from a figure back to the exact source document and location it was read from.

Strix Methodology v1.2 · companion to Kestrel Methodology 4.0. This document describes methodology, which is stable; current coverage figures are maintained live in Kestrel. Strix structures publicly disclosed information; it does not audit or assure it.