AWE

AIR · WATER · EARTH

COLLECTED PROJECT DOCUMENTATION

Environmental Evidence Infrastructure

Research, methodology, architecture and governance

19 documents · Consolidated reading edition
28 September 2026

READING GUIDE

Contents

01	AWE — BRSR Source Landscape	3
02	BRSR Environmental Field Dictionary — first edition	11
03	AWE canonical schema recommendation	15
04	Evidence and provenance contract	21
05	TRS Environmental Evidence Methodology	23
06	AWE Sector Scoring Framework	27
07	Cement Sector Environmental Indicator Dictionary	31
08	AWE Ingestion Architecture & Pipeline Specification	42
09	AWE Canonical Database Schema Specification	46
10	AWE Automated Data Quality Rules & Validation Engine	49
11	AWE Public & Research API Specification (v1.0)	54
12	AWE Dual-Index Search & Evidence-Grounded RAG Architecture	59
13	AWE Public Visualization Library Specification	62
14	AWE Company Profile Specification & UI Wireframe	66
15	AWE Sector Dashboard Specification & UI Wireframe	69
16	AWE Automated Research Story Generator & Templates	71
17	AWE Methodology Versioning & Reproducibility Framework	75
18	AWE Public Governance & Scientific Integrity Framework	77
19	AWE cement MVP — implementation recommendation	80

Compiled from all 19 Markdown files in the docs folder, in filename order. Source wording, status labels, examples and qualifications are retained. This is a reading edition, not a new validation of the claims or a reconciliation of differences between documents. Wide tables are presented as labelled records where helpful. The separate SQL file is outside this Markdown compilation.

DOCUMENT 01

AWE — BRSR Source Landscape

01-source-landscape.md

Research date: 24 September 2026. Version: research-0.1. This is a source audit, not a claim that the historical corpus is complete.

Evidence status vocabulary

- KNOWN FROM SOURCE: directly observed in an identified primary source.
- DERIVED: reproducible result computed from retained source files.
- INFERRED: interpretation or proposed design, not a disclosed fact.
- MISSING: absent from the collected research corpus; not proof of non-existence.
- NEEDS VERIFICATION: located or plausible, but not adequately checked.

Primary source map

SEBI original BRSR circular

URL	https://www.sebi.gov.in/legal/circulars/may-2021/business-responsibility-and-sustainability-reporting-by-listed-entities_50096.html
Format / structure	HTML + PDF; regulatory text
Coverage and years	Voluntary FY2021–22; mandatory top 1,000 FY2022–23 onward
Updates / identifiers	Event driven; circular 2021/562
Reliability and automation	Authoritative rules; not a filing inventory
Reuse / limitations	Public access does not establish a redistribution licence

SEBI Core and revised BRSR

URL	https://www.sebi.gov.in/legal/circulars/jul-2023/brsr-core-framework-for-assurance-and-esg-disclosures-for-value-chain_73854.html
Format / structure	PDF circular and annexures
Coverage and years	Revised format from FY2023–24
Updates / identifiers	Circular 2023/122
Reliability and automation	Core cross-references support programmatic mapping after semantic review
Reuse / limitations	Later amendments supersede portions

SEBI Core definition

URL	https://www.sebi.gov.in/sebi_data/commondocs/jul-2023/Annexure_I-Format-of-BRSR-Core_p.pdf
Format / structure	PDF tables
Coverage and years	Nine ESG attributes; four environmental groups
Updates / identifiers	Regulatory revision
Reliability and automation	Retained raw file; use KPI definitions, not keyword-only matching
Reuse / limitations	Not every environmental BRSR indicator is Core

SEBI industry standards

URL	https://www.sebi.gov.in/legal/circulars/dec-2024/industry-standards-on-reporting-of-brsr-core_90091.html
Format / structure	HTML + PDF
Coverage and years	FY2024–25 onward; applicability text to verify in retained circular
Updates / identifiers	Circular 2024/177
Reliability and automation	Authoritative reporting requirements
Reuse / limitations	Standards version must accompany metric definitions

SEBI assessment / value-chain amendment

URL	https://www.sebi.gov.in/legal/circulars/mar-2025/measures-to-facilitate-ease-of-doing-business-with-respect-to-framework-for-assurance-or-assessment-esg-disclosures-for-value-chain-and-introduction-of-voluntary-disclosure-on-green-credits_93102.html
Format / structure	HTML + PDF
Coverage and years	Amendment issued March 2025
Updates / identifiers	Circular 2025/42
Reliability and automation	Retained raw PDF; updates earlier assurance/value-chain provisions
Reuse / limitations	Do not apply the original 2023 value-chain timetable without this amendment

SEBI assurance FAQ

URL	https://www.sebi.gov.in/sebi_data/faqfiles/apr-2025/1745399101865.pdf
Format / structure	PDF
Coverage and years	April 2025
Updates / identifiers	FAQ version/date
Reliability and automation	Provider independence and standards guidance
Reuse / limitations	Assurance and assessment must remain distinct

NSE taxonomy catalogue

URL	https://www.nseindia.com/static/companies-listing/xbrl-information
Format / structure	HTML; downloadable ZIP, XSD, XML linkbases, XLSX
Coverage and years	Current BRSR entry point 2026-02-28; archive includes 2021-09-30, 2023-06-30, 2024-04-30, 2025-05-31
Updates / identifiers	Event driven; namespaces, QNames, entry points, file hashes
Reliability and automation	Actual packages downloaded and inspected
Reuse / limitations	Shared core schema includes non-BRSR concepts; namespace date is not automatically a reporting-year mandate

NSE current taxonomy package

URL	https://nsearchives.nseindia.com/web/mediaattachment/2026-03/Taxonomy_BRSR_20260330194931.zip
Format / structure	Structured ZIP
Coverage and years	2026-02-28
Updates / identifiers	Entry point and SHA-256
Reliability and automation	KNOWN FROM SOURCE: 12 ZIP entries including reference workbook
Reuse / limitations	Retain exact bytes; publication path date differs from taxonomy version

NSE historical taxonomy package

URL	https://nsearchives.nseindia.com/web/mediaattachment/2026-07/Taxonomy_Archives_20260731164437.zip
Format / structure	Nested ZIPs
Coverage and years	Four archived BRSR versions
Updates / identifiers	Entry points and SHA-256
Reliability and automation	Downloaded; BRSR.zip isolated for local comparison
Reuse / limitations	Not proof of every intermediate release

NSE BRSR repository

URL	https://www.nseindia.com/companies-listing/corporate-filings-bussiness-sustainabilitiy-reports
Format / structure	Dynamic HTML table; PDF/XML attachments; listing CSV; RSS
Coverage and years	Historical range under investigation; live FY2025–26 XML links observed
Updates / identifiers	Company, from/to year, original submission, latest revision, attachment URL
Reliability and automation	Strong filing provenance, company-reported content; interactive retrieval feasible
Reuse / limitations	NSE terms prohibit systematic automated collection. A website endpoint is not a licensed public API. CSV listing is not a metric dataset

BSE BRSR repository

URL	https://www.bseindia.com/corporates/BRSR.aspx
Format / structure	Dynamic filing UI
Coverage and years	NEEDS VERIFICATION of historical XML completeness
Updates / identifiers	BSE code, filing metadata
Reliability and automation	SEBI identifies BSE as an official repository; web request returned 403
Reuse / limitations	API, bulk rights, retention of revisions and terms need verification

BSE filing FAQ compilation

URL	https://www.bseindia.com/markets/MarketInfo/DownloadAttach.aspx?attachedId=2b72ecc3-5dad-4bf7-99f8-54f489fe1a60&id=20250523-40
Format / structure	PDF
Coverage and years	May 2025 compilation; underlying FAQ date to verify
Updates / identifiers	Submission/revision instructions
Reliability and automation	Primary exchange guidance; filing requirements and revision process
Reuse / limitations	Public file; compilation includes internal classification labels; reuse needs review

JK Cement

URL	https://www.jkcement.com/annual-report/ and https://www.jkcement.com/notice-report/
Format / structure	PDF reports, HTML index
Coverage and years	Annual reports listed through FY2025–26; dedicated BRSR FY2023–24 and FY2024–25
Updates / identifiers	Annual; company/report period
Reliability and automation	Primary company evidence, not independently verified merely by publication
Reuse / limitations	Company-site reuse rights to establish; report and metric boundaries vary

UltraTech Cement

URL	https://www.ultratechcement.com/corporate/investors-/financials- and https://www.ultratechcement.com/corporate/sustainability/sustainability
Format / structure	PDF reports, HTML index
Coverage and years	FY2025–26 annual report listed; sustainability archive reaches 2008
Updates / identifiers	Annual; issuer/report/year
Reliability and automation	Historical supplementary evidence; older sustainability reports are not BRSRs
Reuse / limitations	Do not splice frameworks or denominators without reconciliation

Ambuja Cements

URL	https://www.ambujacement.com/annual-report-2025/highlights-of-the-year.html
Format / structure	HTML report
Coverage and years	Current page describes FY2025–26 and prior periods
Updates / identifiers	Annual; report period
Reliability and automation	Direct statement identifies 15-month FY2022–23
Reuse / limitations	URL year is not reliable report-year identification; metric periods need their own check

ACC

URL	https://www.acclimited.com/investors/investor-downloads
Format / structure	PDF + digital BRSR HTML
Coverage and years	FY2025–26 BRSR directly listed; historical archive to inspect
Updates / identifiers	Annual; report/year
Reliability and automation	Digital HTML is parseable but not necessarily XBRL
Reuse / limitations	Inspect reporting boundaries and group overlap with Ambuja

Shree Cement

URL	https://www.shreecement.com/investors/financials-results and https://www.shreecement.com/sustainability/sustainability-reports
Format / structure	PDF and dynamic indexes
Coverage and years	FY2023–24 BRSR and assurance appendix located; broader years to verify
Updates / identifiers	Annual; company/facility/report
Reliability and automation	Also publishes facility environmental reports and CDP responses
Reuse / limitations	Public CDP response does not imply rights over CDP's full dataset

Dalmia Bharat Limited	
URL	https://www.dalmiacement.com/assets/pdf/annual-reports/Dalmia-Bharat-IR-FY-24-25.pdf
Format / structure	PDF
Coverage and years	FY2024–25 BRSR confirmed
Updates / identifiers	Annual; CIN L14200TN2013PLC112346
Reliability and automation	BRSR states consolidated boundary and TUV India reasonable assurance
Reuse / limitations	Distinguish listed Dalmia Bharat Limited from Dalmia Cement (Bharat) and Dalmia Bharat Sugar

BSE FAQ source URL: https://www.bseindia.com/markets/MarketInfo/DownloadAttach.aspx?attachedId=2b72ecc3-5dad-4bf7-99f8-54f489fe1a60&id=20250523-40 . The FAQ describes PDF/XBRL submission and revised filings with a covering letter. Revision capability does not establish public availability of all superseded versions.

Access and reuse findings

KNOWN FROM SOURCE: NSE terms prohibit systematic or automated collection and impose reuse restrictions. NSE data policy describes agreement-based data access. Its applicability and product coverage for BRSR require confirmation. No documented, openly licensed bulk BRSR API has been verified in this investigation. Exchange-to-exchange single filing APIs must not be confused with public research APIs.

INFERRED recommendation: use a permitted exchange feed or licensed delivery for scaled collection. For the research slice, retain individually obtained official downloads with a rights register. Obtain clarity on publication of extracted facts, excerpts, source copies, and commercial redistribution separately. No access-control bypass is part of this plan.

Historical interpretation

Mandatory years are not an availability census. A prior-year value inside a later filing is not a separate prior-year filing. FY2021–22 voluntary filings may contain FY2020–21 comparatives; these must be stored as facts as reported in the later document. FY2026–27 is ongoing at this research date and should not be presented as a completed annual cohort.

Publication timestamps, reporting dates, retrieval timestamps, taxonomy dates, and restatement dates are distinct. Preserve original submission and latest revision metadata without deleting earlier evidence. Period, scope, geography, consolidated/standalone status, acquisitions, and production denominator all govern comparability.

Later source families — discovery backlog, not acquired datasets

CPCB / SPCBs: monitoring and compliance records; MoEFCC / PARIVESH: clearances; CGWB: groundwater; government open-data catalogues: dataset-specific licences; WRI Aqueduct: water-stress context; SBTi: target validation records; scientific earth-observation sources: spatial evidence. Before adoption verify exact dataset, licence, spatial/temporal resolution, facility matching and attribution limits. Ambient measurements and satellite signals cannot automatically be attributed to a company.

DOCUMENT 02

BRSR Environmental Field Dictionary — first edition

02-environmental-field-dictionary.md

Status: DERIVED structural census; INFERRED environmental classification; NEEDS VERIFICATION semantic approval. Source: retained NSE taxonomy packages identified in 01-source-landscape.md. Research date: 2026-09-24.

The complete BRSR-linked concept inventory is retained, including social and governance concepts. The environmental view is a filter over that inventory, not a hand-picked starting list.

Files

- `research/derived/brsr-field-dictionary.csv`: current complete BRSR-linked inventory and requested AWE columns.
- `research/derived/environmental-field-dictionary.csv`: provisional environmental and supporting evidence/context concepts, including structural concepts.
- `research/derived/environmental-core-crosswalk.csv`: 45 explicitly mapped environmental Core-related fact concepts; mapping is an interpretation of the regulatory KPI definitions, not a taxonomy-provided boolean.
- `research/derived/brsr-all-versions.csv`: inventories for all five retained versions.
- `research/derived/taxonomy-changes.csv`: added, removed and changed concept metadata between adjacent versions.
- `research/derived/presentation-relationships.csv`: preserved presentation hierarchy, roles and ordering.
- `research/derived/taxonomy-summary.json`: reproducible census totals.

What is counted

A concept is a distinct schema element ID referenced by the BRSR presentation, definition or calculation linkbases. Shared schema elements not referenced by those BRSR linkbases are excluded. Abstract nodes, axes, tables and domain/member nodes are classified separately from fact concepts. Facts include numeric, text, boolean and date fields: these are not all quantitative environmental indicators. A repeated table concept can yield many observations when combined with dimensions and periods. Counts do not multiply current/prior year cells.

2021-09-30

Shared schema elements	2,013
BRSR-linked concepts	897
Fact concepts	491
Structural concepts	406

2023-06-30

Shared schema elements	2,053
BRSR-linked concepts	938
Fact concepts	528
Structural concepts	410

2024-04-30

Shared schema elements	2,619
BRSR-linked concepts	1,474
Fact concepts	1,043
Structural concepts	431

2025-05-31

Shared schema elements	2,625
BRSR-linked concepts	1,480
Fact concepts	1,048
Structural concepts	432

2026-02-28

Shared schema elements	2,625
BRSR-linked concepts	1,484
Fact concepts	1,052
Structural concepts	432

These are DERIVED counts under the above counting rule, not SEBI's count of questions or mandatory KPIs. The supplied workbook and shared XSD have different inventory scopes; the entry-point linkbases govern this census.

First environmental classification

Current fact concepts: AIR 38; WATER 54; EARTH 47; CROSS-CUTTING 213; NON-AWE 700. Thus 139 fact concepts are provisionally assigned directly to Air/Water/Earth, with 213 supporting or cross-cutting concepts. These are **candidate field counts**, not a reviewed inventory of 352 environmental performance metrics. Cross-cutting includes identity, denominators, applicability, narrative, assurance and shared context. It must not become a performance reward.

Classification uses BRSR presentation roles and explicit name rules. This is reproducible but provisional. Detailed P2 tables and ambiguous generic fields require a semantic review; the retained complete inventory prevents them being lost. `definition` currently contains the official taxonomy label, explicitly identified as a label rather than a complete scientific definition. Instance units, custom denominators, boundary and methodology cannot honestly be filled in from a concept name alone, so unresolved columns say NEEDS VERIFICATION or INSTANCE_UNIT_REQUIRED.

Core interpretation

SEBI Core Annexure I identifies four environmental attribute groups: GHG footprint (P6 E7), water footprint (P6 E3/E4), energy footprint (P6 E1), and circularity/waste (P6 E9). Core is not equivalent to all Principle 6 fields. Scope 3, non-GHG pollutants, biodiversity narratives and water-stress facility tables are not automatically Core KPIs. Water withdrawal should not be relabelled Core water consumption.

The 45-concept crosswalk includes totals, intensity variants and discharge/waste breakdown concepts. It is not a claim that there are exactly 45 environmental regulatory questions. Renewable energy share needs an explicit derivation from compatible total renewable energy and total energy facts; it must not be substituted with renewable electricity share. Some concepts support Core measurement without independently being a Core KPI. Assurance metadata never proves that a particular fact was covered by the assurance engagement.

Changes actually observed

Transition	Added linked concepts	Removed	Metadata changed
2021 → 2023	42	1	37
2023 → 2024	586	50	884
2024 → 2025	19	13	462
2025 → 2026	4	0	0

2023 adds notes and applicability explanations. The 2024 package expands assurance metadata and revised environmental intensity/energy fields. The 2025 package adds green-credit fields and changes assessment/assurance provider concepts. In the inspected 2026 BRSR-linked inventory, the four additions are ISIN, MSEISymbol, NSESymbol and ScripCode; they already existed in the shared schema. This is an entry-point linkage change, not four new environmental measures.

Metadata changes include references and presentation ancestry. They do not imply that 884 underlying measurement definitions changed. The comparison does not yet prove equivalence of all dimensional restrictions, calculation arcs, custom type enumerations or validation rules. Never align versions solely by

local tag name, nor infer the filing's taxonomy from its FY label. Read its schemaRef.

Review gates

1. Review every candidate's presentation path and regulatory question, including environmental content in P2, Section B and narrative fields elsewhere.
2. Resolve custom type enumerations and any fields classified as structural by naming convention.
3. Verify official industry standards and full Core-to-concept scope, including calculation inputs and dimensional breakdowns.
4. Cross-check actual instances against taxonomy and report tables; preserve errors and original units.
5. Publish a versioned reviewed crosswalk only after these checks. The first edition is useful for ingestion research, not for publishing ratings.

DOCUMENT 03

AWE canonical schema recommendation

03-canonical-schema.md

Status: INFERRED design proposal grounded in the inspected taxonomies and JK Cement instance. No production database or public rating has been implemented.

Design decision

Use PostgreSQL as the relational evidence graph, object storage for immutable bytes, and Python for deterministic parsing. An observation is an assertion made by a source, not a single mutable truth cell. Several observations may describe the same company, metric and period. Reconciliation produces a separate, versioned selection decision; it does not delete disagreements.

Company identifiers must be time-aware. ISIN belongs to a security, not necessarily uniquely to a company. NSE symbols and BSE codes identify listings. Corporate succession and group ownership need effective dates. Do not double-count ACC inside an Ambuja consolidated total when building sector aggregates.

Relations

Entity	Key and important fields	Relationships / invariants
company	company_id UUID; legal_name; incorporation_date; jurisdiction	Stable entity identity; aliases and identifiers in child tables
company_identifier	identifier_id; company_id; scheme; value; valid_from/to; evidence_id	CIN/LEI etc.; preserve identifier history; no inferred mergers
company_alias	company_id; name; valid_from/to; evidence_id	Renames do not create a new environmental history automatically
security	security_id; company_id; ISIN; class; valid_from/to	Multiple securities per company
listing	listing_id; security_id; exchange; symbol; exchange_code; valid_from/to	Time-aware NSE/BSE mappings
sector	sector_id; classification_system; version; code; parent_id	Versioned company_sector assignments; source and dates
company_relationship	parent_id; child_id; relationship_type; ownership; valid_from/to; evidence_id	Subsidiaries, mergers, acquisitions, divestments
facility	facility_id; name; address; coordinates; geography_precision	facility_ownership dates; do not assign environmental events by name similarity alone
source	source_id; authority; source_type; base_url; reliability_category; rights_policy_id	Separate publisher from underlying reporting company
rights_policy	policy_id; source_url; captured_at; permitted_uses; restrictions; review_status	Raw copy, excerpts, extracted facts and commercial redistribution are distinct uses
raw_object	SHA256 primary key; storage_key; byte_length; media_type; first_seen_at	Write-once bytes; storage credentials cannot replace existing objects

Entity	Key and important fields	Relationships / invariants
retrieval	retrieval_id; source_id; requested_url; final_url; retrieved_at; HTTP metadata; raw_hash	A URL can change bytes; one object can appear at several URLs
report	report_id; company_id; report_type; financial_year_label; publication_date	Logical report, not its file
report_version	report_version_id; report_id; raw_hash; taxonomy_version_id; exchange_filing_id; submitted_at; supersedes_id; amendment_reason	Original/revised versions retained; revision known/unknown rather than default false
reporting_period	period_id; start_date; end_date; instant_date; duration_days; label	Duration vs instant mutually exclusive; dates determine comparability
reporting_boundary	boundary_id; basis; consolidation_method; geography; scope_definition; evidence_id	Membership in boundary_member with company/facility and effective dates; metric-specific exceptions
taxonomy_version	taxonomy_version_id; namespace; entry_point; package_hash; publication_date; effective_rule_evidence_id	Never assume taxonomy date equals fiscal-year applicability
taxonomy_concept	taxonomy_version_id + QName; original_label; datatype; period_type; abstract; source_location	Retain all ESG concepts; hierarchy/dimensions/reference arcs in related tables
concept_metric_mapping	mapping_id; concept_id; metric_id; mapping_version; dimensions_rule; status; reviewer	Many-to-many mappings with effective taxonomy versions; explicit exclusions
metric	metric_id; canonical_name; pillar; topic; quantity_kind; direction; definition_version	Metric definition includes gross/net emissions, scope and denominator semantics
unit	unit_id; symbol; dimension; canonical_unit; exact_factor; offset; conversion_version	Units with monetary/physical denominators are compound; store components separately
observation	observation_id; company_id; report_version_id; concept_id; context_id; period_id; boundary_id; metric_id nullable; original_lexical_value; original_numeric_value NUMERIC nullable; original_unit_id; value_state; source_location; original_text; extraction_run_id	Append-only; retain null/nil/not-applicable/blank/dash/zero separately; unmapped facts allowed
xbrl_context	context_id; report_version_id; original_context_id; entity_scheme/value; period_id; dimensions_json; original_xml	Explicit and typed dimensions preserved; context IDs are document-local
xbrl_unit	report_version_id + original_unit_id; numerator_measures; denominator_measures; original_xml	Resolve QName namespaces, not just strings such as 'pure' or 'Tonne'
normalized_observation	normalization_id; observation_id; normalized_value; unit_id; factor; rule_version; run_id; validation_status	New result for new rule; never overwrite source assertion
derived_result	derived_id; metric_id; result_value; unit_id; run_id; period_id; boundary_id	derived_input links every operand, formula and denominator, including PPP vintage
evidence	evidence_id; raw_hash; report_version_id; locator_type; page/table/bbox/XPath; quote; quote_hash	One evidence object may support or contradict several assertions
claim	claim_id; company_id; claim_type; original_text; date; scope; alignment_status; status_version	claim_evidence join stores support/contradiction/context; no automatic greenwashing label

Entity	Key and important fields	Relationships / invariants
target	target_id; claim_id; metric_id; baseline_observation_id; target_date; target_value; unit; target_type; absolute_or_intensity; boundary_id; scope; validation_status	target_version and target_evidence; SBTi status dated and sourced, not an unqualified boolean
assurance	assurance_id; provider_entity; engagement_type; assurance_level; standard; signed_date; opinion; exclusions; evidence_id	assurance_scope joins covered metrics/observations/periods/boundaries; report-level Y is insufficient
contradiction	contradiction_id; classification; materiality; status; created_run_id	contradiction_observation joins competing assertions; resolution is separately logged
regulatory_event	event_id; authority; docket_id; event_type; status; date; facility_id; evidence_id	Allegation, finding, appeal and final decision distinguished; no unsourced adverse inference
quality_issue	issue_id; observation_id; rule_version; severity; message; resolution_id	VALID/WARNING/ERROR/MANUAL_REVIEW; source values never silently fixed
benchmark	benchmark_id; metric_id; sector_id; period_id; boundary_rule; denominator_rule; dataset_snapshot_id; sample_n; median; quartiles; method_version	benchmark_member enumerates actual included observations and exclusions
methodology_version	methodology_id; version; status; effective_date; config_hash; code_revision; published_document_hash	Draft/reviewed/published; immutable once published
calculation_run	run_id; methodology_id; code_revision; dependency_lock_hash; input_snapshot_id; executed_at; parameters_hash	Every normalization, derivation, benchmark and score reproducible
awe_score	score_id; company_id; period_id; methodology_id; run_id; pillar_values; performance; coverage; publication_status	Separate score_indicator rows with expected weights and usable inputs; no disclosure reward
evidence_confidence	confidence_id; assertion_id; model_version; component_values; score nullable; limitations	Separate from performance; extraction confidence is a separate measurement again
correction / review_decision	decision_id; subject_id; reason; old_version; new_version; reviewer; evidence_id; timestamp	Public material correction log; preserve both old and new calculations

Observation identity and idempotency

Raw deduplication uses SHA-256, with retrieval events stored separately. Parsed fact identity uses document hash + stable source locator + parser version. A fact's semantic comparison key uses company + metric definition + exact period + canonical dimension signature + boundary + unit/denominator semantics. Do not deduplicate on company/metric/year: that would destroy restatements and contradictory sources.

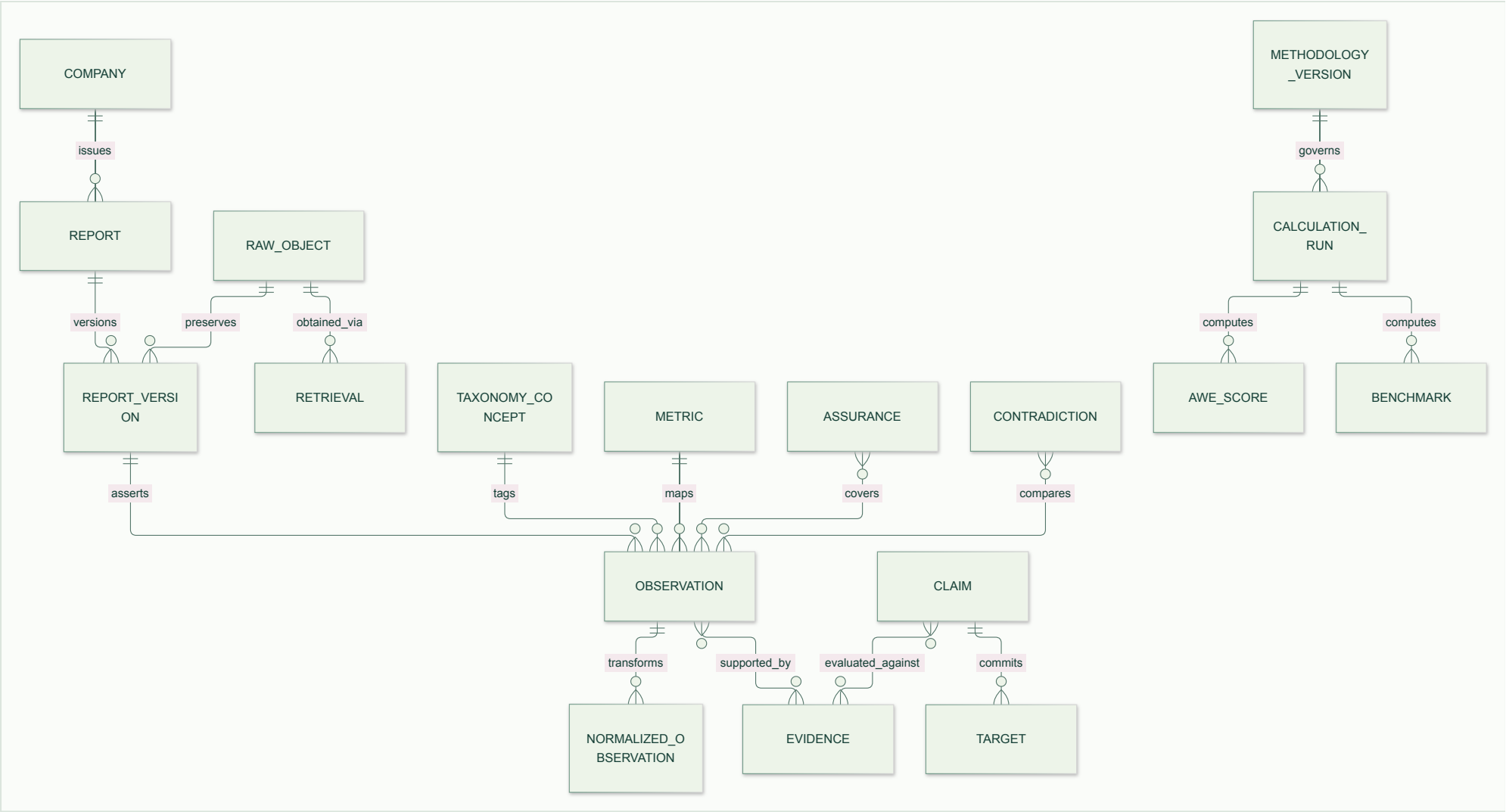
Use a unique constraint on the parse-run/document/locator combination. Canonical mappings and normalizations have their own version keys. A repeated download of identical bytes must not create new facts; a new parser run may create a new derived parse version while preserving the previous result. Cross-exchange identical hashes can share an object; distinct files remain distinct even if some facts agree.

Comparability and validation

A comparability decision records HIGH/MEDIUM/LOW/NOT_COMPARABLE, reasons, reviewer/rule version, and the two observations compared. Exact period length, production boundary, energy source, Scope 2 accounting basis, gross/net emissions, product mix and denominator must be compatible. Financial consolidation alone is insufficient.

Decimal arithmetic is required. Negative environmental quantities trigger review, not silent clamping. A percentage rule applies only to concepts whose definition limits the range to 0–100. Do not rescale a 'pure' XBRL fact from 0–1 based on its magnitude. Water balances are conditional on method, storage and boundary. PPP intensity must retain exchange factor, reference year and units; revenue-normalised and output-normalised intensities are different metrics.

Evidence graph


COMPANY
issues
REPORT
RAW_OBJECT
versions
preserves
obtained_via
REPORT_VERSION
RETRIEVAL
asserts
TAXONOMY_CONCEPT
tags
METRIC
maps
ASSURANCE
covers
CONTRADICTION
compares
METHODOLOGY_VERSION
governs
CALCULATION_RUN
computes
computes
AWE_SCORE
BENCHMARK
OBSERVATION
CLAIM
transforms
supported_by
evaluated_against
commits
NORMALIZED_OBSERVATION
EVIDENCE
TARGET

Performance, evidence and coverage

Keep three independent outputs. Performance consumes only eligible environmental outcomes, trajectories and target results. Confidence consumes provenance and verification evidence. Coverage is usable materiality weight / expected materiality weight. Missing facts do not become zero performance; avoid renormalising scores across radically different observed subsets and presenting them as comparable. Suppress composite ratings until minimum pillar and overall coverage gates are approved and met.

The supplied TRS weights are a research hypothesis, not a validated probability model. Store all components, definitions and review decisions; do not label the resulting weighted score a calibrated probability of truth. No separate TrustRankScore source methodology was supplied, so its lineage remains NEEDS VERIFICATION. Contradiction handling and assurance scope must be explicit rather than hidden penalties.

DOCUMENT 04

Evidence and provenance contract

04-evidence-provenance.md

Status: INFERRED specification; first raw taxonomy packages and a real company XBRL instance retained locally.

Required chain

SOURCE → RAW OBJECT → REPORT VERSION → ORIGINAL FACT → NORMALIZATION → DERIVATION → BENCHMARK → SCORE → PUBLICATION.

Every hop has a stable ID and version. Evidence confidence and extraction confidence remain separate from environmental performance.

Raw evidence

Store bytes under content-derived keys such as `raw/sha256/ab/<hash>`. Use conditional create and verify size/hash. A retrieval manifest stores URL, redirects, retrieval timestamp, content type, status, source authority, rights status and SHA-256. Repeated bytes reuse the same object and add a retrieval event. Local research files are retained unchanged; production object-lock and permission enforcement remain to implement. Hashes detect change but do not themselves enforce immutability or establish truth.

XBRL

Retain schemaRef, concept QName and namespace; XML fact ordinal/XPath; lexical value; contextRef; unitRef; decimals/precision; xsi:nil; fact ID if present; full context and unit XML. Resolve duration/instant periods, entity identifiers, explicit dimensions, typed dimensions and divided units. Two facts with identical concept and period but different dimensions are not duplicates. Duplicate same-context facts with different values become quality issues.

Store all facts before semantic filtering. Validate schemaRef against retained taxonomy packages. An unrecognised namespace is quarantined for review; do not silently map it to the newest schema. External XML entities and automatic network schema fetching should be disabled. Record taxonomy validation separately from arithmetic and environmental plausibility checks.

PDF and narrative evidence

Preserve document hash, physical PDF page index, printed page label, table/section, bounding box where available, original snippet, extractor name/version and confidence. A chart point must resolve to an observation and then to a stable source location. A URL alone is insufficient.

AI extraction outputs extracted_value (nullable), source_quote, source_location, confidence, reasoning_category and needs_human_review. Store model/prompt versions and structured extraction schema. Unsupported facts remain absent; never fill missing values from model memory. Permission to retain full source bytes is distinct from permission to republish excerpts.

Assurance

Represent assessment, limited assurance and reasonable assurance separately. Capture provider identity, engagement standard, period, organizational boundary, included indicators, exclusions, opinion and signed statement. A company checkbox is a reported assertion about assurance, not independent proof. Provider declarations in XBRL are useful discovery keys; verify the engagement statement before attaching coverage to metrics. Assurance of financial statements does not automatically cover BRSR facts.

Restatements and reconciliation

Preserve each filing and each current/prior-year assertion. If a FY2025–26 report changes its FY2024–25 comparative, retain both the original FY2024–25 disclosure and the later comparative. Create a relationship of restates or conflicts-with supported by evidence. A change without a restatement explanation is a review item, not an automatic correction. A default public view may choose an approved latest comparable assertion, while offering as-reported and restated views.

Published value payload

At minimum: observation/derived-result ID, company, metric definition version, exact period, original value and unit, normalized value and unit if any, reporting boundary, dimensions, report version, source URL, SHA-256, locator, reported/verified/derived/estimated/modeled status, transformation formula/version, input IDs, assurance scope, quality flags, comparability flags and methodology version. Never present extraction confidence as evidence confidence.

Correction lifecycle

Proposed correction → evidence review → approved new version → recalculation with pinned inputs → changelog → publication. Companies may submit factual corrections; commercial relationships cannot change scoring logic. Keep reviewer identity, conflict-of-interest declaration and reasons. No deletion of inconvenient evidence or historical methodology outcomes.

DOCUMENT 05

TRS Environmental Evidence Methodology

05-trs-evidence-methodology.md

Document ID: AWE-TRS-METH-v1.0
Status: INFERRED methodology specification grounded in SEBI BRSR Core, LODR regulations, and XBRL filings.
Applicability: Environmental evidence assessment across Air, Water, Earth pillars.

---

1. Philosophical Grounding: Evidence vs. Performance

AWE enforces a strict operational separation between two orthogonal vectors:

1. **Environmental Performance (P):** What is the physical impact of the company's activities on air, water, and soil systems?
2. **Evidence Confidence (E):** How verifiable, accountable, rigorous, and methodologically transparent is the evidence underpinning that reported performance?

> [!CRITICAL] **Cardinal Principle:** Better reporting quality or extensive third-party assurance must **never** inflate environmental performance. A high-emissions, heavily polluting cement plant with reasonable assurance from a Big Four firm must receive a low Performance score combined with a high Evidence Confidence score.

---

2. TRS Environmental Evidence Hierarchy (E0 to E5)

AWE classifies every observation, assertion, and claim into an empirical six-tier evidentiary hierarchy:

E5

Name	**Statutory Filing with Independent Assurance**
Description	Mandated regulatory disclosure submitted under statutory oath (SEBI LODR) subject to independent assurance under ISAE 3000 / AA1000AS or SEBI-mandated BRSR Core Reasonable/Limited Assurance.
Base Confidence Anchor	**90% – 100%**
Typical Examples in Indian Corpus	BRSR Core Principle 6 GHG / Water data with attached signed Independent Assurance Report by TUV India, SGS, DNV, or BDO.

E4

Name	**Statutory Regulatory Disclosure**
Description	Mandated regulatory filing submitted to stock exchanges under statutory liability (e.g. SEBI BRSR XBRL / PDF) without explicit indicator-level assurance.
Base Confidence Anchor	**70% – 85%**
Typical Examples in Indian Corpus	BRSR Section C Principle 6 non-core indicators (e.g., Scope 3 emissions, biodiversity narrative) signed by Company Secretary / Managing Director.

E3

Name	**Assured Voluntary Disclosure**
Description	Voluntary sustainability report, GRI report, or CDP filing backed by third-party assurance, but outside direct statutory exchange liability.
Base Confidence Anchor	**55% – 70%**
Typical Examples in Indian Corpus	Annual Integrated Sustainability Report assured by external auditor; verified CDP Climate Change questionnaire.

E2

Name	**Disclosed Company Methodology**
Description	Disclosed metric with explicit formula, emission factors, boundaries, and primary metering, but unverified by an independent auditor.
Base Confidence Anchor	**40% – 55%**
Typical Examples in Indian Corpus	Company Investor Presentation or ESG Factsheet detailing specific clinker factor calculations and metering methodology.

E1

Name	**Uncorroborated Corporate Narrative**
Description	Self-reported narrative claim, PR release, marketing text, or high-level CSR claim lacking baseline, denominator, or raw figures.
Base Confidence Anchor	**15% – 30%**
Typical Examples in Indian Corpus	"We are committed to becoming green and water positive by 2030" on company website without annual water balance.

E0	
Name	**Unsupported / Conflicting Assertion**
Description	Claim refuted by regulatory findings, contradictory across sources without explanation, or mathematically impossible.
Base Confidence Anchor	**0% – 10%**
Typical Examples in Indian Corpus	Water withdrawal reported as zero for an operational wet/dry cement plant; net-zero claim with zero Scope 1 reduction.

3. Mathematical Scoring Formulation

For any environmental observation or claim i, the Evidence Confidence score $EC_i \in [0, 100]$ is computed deterministically:

$$EC_i = w_{SQ} \cdot SQ_i + w_{IA} \cdot IA_i + w_{SQN} \cdot SQN_i + w_{MT} \cdot MT_i + w_{REC} \cdot REC_i + w_{ECR} \cdot ECR_i - \Delta_{PEN}$$

3.1 Candidate Dimension Weights (Hypothesis v1.0)

Dimension	Symbol	Weight (w_k)	Core Evaluation Criteria
Source Quality	SQ	**25%**	Legal provenance, statutory liability, regulatory enforcement, publisher authority (SEBI/NSE/BSE = 100; Company statutory report = 85; Standalone CSR brochure = 40; Press release = 20).
Independent Assurance	IA	**20%**	Independent assurance level (Reasonable = 100; Limited = 75; Management assessment = 35; None = 0); auditor independence, accredited standard (ISAE 3000 / ISAE 3410 / AA1000AS). Scope must cover the specific metric!
Specificity & Quantification	SQN	**20%**	Numerical precision, standard dimensional units (tCO2e, MWh, m³, MT), clear denominator, geographic and facility boundary disclosure.
Methodology Transparency	MT	**15%**	Explicit emission factors (IPCC / CEA CO2 Baseline Database for Indian Power Sector), calculation standards (GHG Protocol, ISO 14064, CPCB guidelines).
Recency	REC	**10%**	Age of disclosure relative to assessment year: Current FY = 100; FY-1 = 85; FY-2 = 65; FY-3 = 40; Older = 20.
External Corroboration	ECR	**10%**	Alignment with secondary primary datasets: CPCB OCEMS monitoring, CEA power generation records, satellite NOx/SO2 observations, SPCB compliance filings.

3.2 Penalties (Δ_{PEN})

- **Unexplained Contradiction:** -20 points if the metric contradicts another regulatory filing for the same period without an explicit restatement note.
- **Reporting Boundary Ambiguity:** -15 points if boundary shifts from Consolidated to Standalone without comparative restatement.
- **Unit Ambiguity:** -15 points if raw filing omits scale (e.g. reporting thousands of tonnes as raw tonnes).

4. Operational Assurance Verification Rules

1. **Indicator-Level vs. Report-Level Assurance:** A company checking "Yes" to general BRSR assurance does **not** grant $IA = 100$ to all 1,484 fields. The assurance statement text or XBRL assurance mapping (`assurance-source-facts.jsonl`) must be parsed to confirm that the specific indicator (e.g. Scope 1 GHG) falls within the approved engagement boundary.
2. **Standard Validation:** The assurance provider must be an independent accredited entity. In our cement census, verified providers include:
   - TUV India Private Limited (JK Cement FY25–26, Dalmia Bharat FY24–26, Ambuja FY25)
   - SGS India Private Limited (Shree Cement FY24, FY26; Ambuja FY26; ACC FY26)
   - BDO India LLP (UltraTech FY24, FY25)
   - TUV SUD South Asia (UltraTech FY26)
   - Intertek India (Ambuja FY24, Shree Cement FY25)
3. **Reasonable vs. Limited Assurance:** Under SEBI Circular `SEBI/HO/CFD/CFD-SEC-2/P/CIR/2023/122`, mandatory assurance of BRSR Core begins with *reasonable assurance* for top listed entities. If only *limited assurance* is provided, IA is scored at 75%. If self-assessed, IA is scored at 35%.

5. Aggregation from Observation to Company Pillar Confidence

For a given company c and environmental pillar $P \in \{\text{AIR}, \text{WATER}, \text{EARTH}\}$:

$$EC_{c,P} = \frac{\sum_{i \in \mathcal{M}_P} \mu_i \cdot \delta_i \cdot EC_i}{\sum_{i \in \mathcal{M}_P} \mu_i}$$

Where:

- $\mathcal{M}_P$ is the set of required indicators in pillar P.
- μ_i is the regulatory/materiality weight of indicator i.
- $\delta_i = 1$ if an observation exists, and 0 if missing. (Note: missing indicators reduce **Score Coverage**, while existing indicators carry their empirical EC_i).

Overall Company Evidence Confidence:

$$EC_c = w_{Air} \cdot EC_{c,Air} + w_{Water} \cdot EC_{c,Water} + w_{Earth} \cdot EC_{c,Earth}$$

DOCUMENT 06

AWE Sector Scoring Framework

06-sector-scoring-framework.md

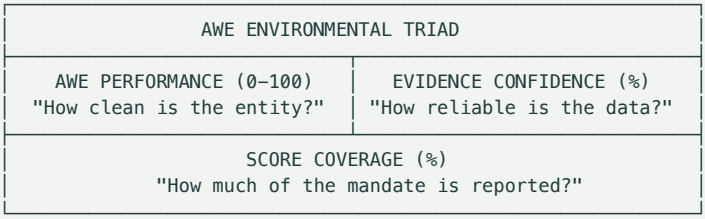
Document ID: AWE-SCORE-FRAME-v1.0
Status: INFERRED methodology specification. Sector weights subject to stakeholder materiality review.
Version Anchor: AWE-SECTOR-2026-v1.0

---

1. Core Architecture and Separation Principles

AWE generates three independent metrics for every evaluated entity:

1. **AWE Performance Score ($S_{AWE} \in [0, 100]$):** Pure physical environmental performance relative to peer benchmarks and scientific decarbonisation trajectories.
2. **Evidence Confidence ($EC \in [0, 100\%]$):** TRS-derived confidence in the veracity and auditability of the underlying disclosures.
3. **Score Coverage ($SC \in [0, 100\%]$):** Completeness of usable observations across expected material indicators.



> [!WARNING] **Anti-Greenwashing Constraint:** Under no circumstances shall missing data be defaulted to zero emissions or median performance. Missing data reduces **Score Coverage** and triggers high uncertainty penalties, but does not allow poor performers to hide behind omission.

---

2. Pillar Decomposition and Sector Materiality Weights

The composite performance score is defined as:

$$S_{AWE} = w_{Air} \cdot S_{Air} + w_{Water} \cdot S_{Water} + w_{Earth} \cdot S_{Earth}$$

Where $\sum_P w_P = 1.0$. Sector weights reflect physical environmental materiality:

Cement

Air (w_{Air})	**45%**
Water (w_{Water})	**25%**
Earth (w_{Earth})	**30%**
Materiality Rationale	Massive process Scope 1 emissions (calcination), thermal energy intensity, quarrying/limestone extraction, and circular kiln co-processing (AFR).

Steel

Air (w_{Air})	**50%**
Water (w_{Water})	**20%**
Earth (w_{Earth})	**30%**
Materiality Rationale	High carbon blast-furnace emissions, slag utilization, scrap recycling.

Power (Thermal)

Air (w_{Air})	**60%**
Water (w_{Water})	**25%**
Earth (w_{Earth})	**15%**
Materiality Rationale	Dominant coal combustion, cooling water consumption, fly-ash management.

Chemicals

Air (w_{Air})	**35%**
Water (w_{Water})	**35%**
Earth (w_{Earth})	**30%**
Materiality Rationale	Toxic air releases, hazardous effluent treatment, hazardous chemical waste.

FMCG / Food & Bev

Air (w_{Air})	**25%**
Water (w_{Water})	**45%**
Earth (w_{Earth})	**30%**
Materiality Rationale	Agricultural water intensity, packaging plastics, EPR compliance.

IT & Financial Services	
Air (w_{Air})	**50%**
Water (w_{Water})	**20%**
Earth (w_{Earth})	**30%**
Materiality Rationale	Data center electricity (Scope 2), e-waste management, minimal direct water.

---

3. Sub-Pillar Structure

Within each pillar $P \in \{\text{AIR}, \text{WATER}, \text{EARTH}\}$, performance is evaluated across four deterministic components:

$$S_P = 0.40 \cdot C_{intensity} + 0.25 \cdot C_{trajectory} + 0.20 \cdot C_{target} + 0.15 \cdot C_{operational}$$

3.1 Current Intensity ($C_{intensity} \in [0, 100]$)

Measures the entity's normalized physical intensity against the peer benchmark distribution:

- For lower-is-better indicators (e.g. Specific GHG emissions per tonne of cement, Specific water consumption per tonne):

  $$I_{norm} = \max\left(0, \min\left(100, 100 \cdot \frac{B_{worst} - X_{current}}{B_{worst} - B_{best}}\right)\right)$$

  Where B_{worst} is the 90th percentile of the sector and B_{best} is the 10th percentile (Winsorized to protect against outliers).

3.2 Historical Trajectory ($C_{trajectory} \in [0, 100]$)

Evaluates 3-year Compound Annual Reduction Rate ($CARR$) or YoY improvement:

$$CARR = 1 - \left(\frac{X_t}{X_{t-k}}\right)^{\frac{1}{k}}$$

- If $CARR \geq CARR_{aligned}$ (e.g., SBTi 1.5°C minimum 4.2% annual reduction): Score = 100.
- If $CARR \leq 0$ (emissions growing): Score = 0 to 30.
- Boundary Shift Override: If a material change in reporting boundary occurred between $t - k$ and t, $C_{trajectory}$ is flagged as `NOT_COMPARABLE` and re-weighted to $C_{intensity}$.

3.3 Target Progress ($C_{target} \in [0, 100]$)

Evaluates verified performance against stated commitments:

- For reduction targets (lower-is-better):

  $$\text{Progress} = \frac{X_{baseline} - X_{current}}{X_{baseline} - X_{target}}$$

- Expected Progress at year t:

$$\text{Progress}_{expected} = \frac{t - t_{baseline}}{t_{target} - t_{baseline}}$$

- Trajectory Ratio (TR):

$$TR = \frac{\text{Progress}}{\text{Progress}_{expected}}$$

  - $TR \geq 1.0$: **ON TRAJECTORY** (Score: 90–100)
  - $0.75 \leq TR < 1.0$: **NEAR TRAJECTORY** (Score: 60–80)
  - $0.0 \leq TR < 0.75$: **OFF TRAJECTORY** (Score: 20–50)
  - $\text{Progress} < 0$: **REGRESSING** (Score: 0)

3.4 Operational & Compliance Outcomes ($C_{operational} \in [0, 100]$)

Evaluates clean technology adoption and environmental compliance:

- Share of renewable electricity / Waste Heat Recovery System (WHRS) generation.
- Thermal Substitution Rate (TSR) from alternative fuels.
- Zero Liquid Discharge (ZLD) status and zero regulatory penalty notices from CPCB/SPCB.

---

4. Score Coverage Formulation

Score Coverage (SC_c) evaluates what fraction of material indicators have empirical, validated observations:

$$SC_c = \frac{\sum_{i \in \mathcal{M}_{sector}} \mu_i \cdot \mathbf{1}_{\{X_{i,c} \text{ is usable}\}}}{\sum_{i \in \mathcal{M}_{sector}} \mu_i}$$

Where μ_i is the materiality weight of indicator i.

Minimum Publication Gates

- **Comprehensive Score:** Requires $SC_c \geq 75\%$ overall and $SC_P \geq 60\%$ in all three pillars.
- **Provisional / Incomplete Rating:** If $50\% \leq SC_c < 75\%$, score is published with an prominent `[INCOMPLETE DATA]` caveat.
- **Suppressed Rating:** If $SC_c < 50\%$, no composite score is generated; only raw observations and a missingness audit are presented.

DOCUMENT 07

Cement Sector Environmental Indicator Dictionary

07-cement-indicator-dictionary.md

Document ID: AWE-DICT-CEMENT-v1.0
Sector Classification: NIC 23941 / 23942 — Manufacture of Clinker and Hydraulic Cement
Target Entities: UltraTech Cement, Ambuja Cements, ACC, Shree Cement, Dalmia Bharat, JK Cement
Research Date: 2026-09-24

---

1. Overview of Cement Industry Materiality

Cement manufacturing is among the world's most energy- and carbon-intensive industrial processes. Process emissions from the calcination of limestone ($CaCO_3 \rightarrow CaO + CO_2$) account for roughly 60% of Scope 1 emissions, while fuel combustion in rotary kilns accounts for the remaining ~40%. Therefore, physical output denominators (Tonnes of Clinker vs. Tonnes of Cementitious Material) are vital to prevent erroneous comparisons.

---

2. Core Indicator Matrix

Pillar I: AIR & CLIMATE (Pillar Weight: 45%)

`AIR-GHG-001`

Canonical Metric Name	**Gross Scope 1 GHG Emissions**
XBRL Concept QName	`in-capmkt_TotalScope1Emissions`
Canonical Unit	tCO_2e
Alternate Reported Units	$MtCO_2e$, Lakh tCO_2e
Normalization Conversion	If reported in Million: $\times 10^6$; If reported in Lakh: $\times 10^5$
Denominator Rule	Absolute
Core Flag	YES

`AIR-GHG-002`

Canonical Metric Name	**Gross Scope 2 GHG Emissions**
XBRL Concept QName	`in-capmkt_TotalScope2Emissions`
Canonical Unit	$\mathrm{tCO_2e}$
Alternate Reported Units	$\mathrm{MtCO_2e}$, $\mathrm{tCO_2e}$
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute (Grid market/location)
Core Flag	YES

`AIR-GHG-003`

Canonical Metric Name	**Net Scope 1 GHG Emissions**
XBRL Concept QName	`Derived / Sustainability Report`
Canonical Unit	$\mathrm{tCO_2e}$
Alternate Reported Units	$\mathrm{tCO_2e}$
Normalization Conversion	Gross Scope 1 minus biogenic fuels
Denominator Rule	Absolute
Core Flag	NO (Leadership)

`AIR-INT-001`

Canonical Metric Name	**GHG Intensity per Turnover**
XBRL Concept QName	`in-capmkt_TotalGreenHouseGasEmissionsPerRupeeOfTurnover`
Canonical Unit	$\mathrm{tCO_2e/INR\ Cr}$
Alternate Reported Units	$\mathrm{tCO_2e/INR}$
Normalization Conversion	Factor: $\times 10^7$ (to convert per INR to per Crore)
Denominator Rule	Turnover (INR Cr)
Core Flag	YES

`AIR-INT-002`

Canonical Metric Name	**Specific Net CO2 per Tonne Cement**
XBRL Concept QName	`in-capmkt_GreenHouseGasEmissionsPerPhysicalOutput`
Canonical Unit	$\mathrm{kgCO_2e/t}$
Alternate Reported Units	$\mathrm{tCO_2e/t}$
Normalization Conversion	If $\mathrm{tCO_2e/t}$: $\times 1000$
Denominator Rule	Tonnes of Cementitious Product
Core Flag	YES

`AIR-ENE-001`

Canonical Metric Name	**Total Energy Consumption**
XBRL Concept QName	`in-capmkt_TotalEnergyConsumptionInJoules`
Canonical Unit	GJ
Alternate Reported Units	MWh, TJ
Normalization Conversion	If MWh: $\times 3.6$; If TJ: $\times 1000$
Denominator Rule	Absolute
Core Flag	YES

`AIR-ENE-002`

Canonical Metric Name	**Renewable Electricity Share**
XBRL Concept QName	`Derived (Renewable MWh / Total MWh)`
Canonical Unit	$\%$
Alternate Reported Units	Pure fraction $(0.0 - 1.0)$
Normalization Conversion	If decimal: $\times 100$
Denominator Rule	Total Electricity Consumed
Core Flag	YES

`AIR-ENE-003`

Canonical Metric Name	**Thermal Substitution Rate (TSR)**
XBRL Concept QName	`Derived (AFR Energy / Kiln Energy)`
Canonical Unit	%
Alternate Reported Units	%
Normalization Conversion	Retain percentage
Denominator Rule	Total Thermal Kiln Energy
Core Flag	NO (Sector Key)

`AIR-ENE-004`

Canonical Metric Name	**Waste Heat Recovery (WHRS) Power**
XBRL Concept QName	`Corporate Sustainability / BRSR narrative`
Canonical Unit	MW / MWh
Alternate Reported Units	MW, MWh
Normalization Conversion	Retain capacity and generation
Denominator Rule	Plant Generation
Core Flag	NO (Sector Key)

`AIR-POL-001`

Canonical Metric Name	**Particulate Matter (PM) Emissions**
XBRL Concept QName	`in-capmkt_DetailsOfAirEmissionsPM`
Canonical Unit	Tonnes
Alternate Reported Units	kg, MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	NO (Essential)

`AIR-POL-002`

Canonical Metric Name	**NOx Emissions**
XBRL Concept QName	`in-capmkt_DetailsOfAirEmissionsNOx`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	NO (Essential)

`AIR-POL-003`

Canonical Metric Name	**SOx Emissions**
XBRL Concept QName	`in-capmkt_DetailsOfAirEmissionsSOx`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	NO (Essential)

Pillar II: WATER RESILIENCE (Pillar Weight: 25%)

`WAT-WIT-001`

Canonical Metric Name	**Total Water Withdrawal**
XBRL Concept QName	`in-capmkt_TotalVolumeOfWaterWithdrawal`
Canonical Unit	m^3 (or kl)
Alternate Reported Units	kl, Million m^3
Normalization Conversion	$1\ \mathrm{kl} = 1\ \mathrm{m}^3$; If Mm^3: $\times 10^6$
Denominator Rule	Absolute
Core Flag	YES

`WAT-WIT-002`

Canonical Metric Name	**Groundwater Withdrawal**
XBRL Concept QName	`in-capmkt_WaterWithdrawalByGroundwater`
Canonical Unit	m^3
Alternate Reported Units	kl
Normalization Conversion	$1\ \mathrm{kl} = 1\ \mathrm{m}^3$
Denominator Rule	Absolute
Core Flag	YES

`WAT-WIT-003`

Canonical Metric Name	**Water-Stressed Withdrawal Share**
XBRL Concept QName	`in-capmkt_TotalVolumeOfWaterWithdrawalPerArea`
Canonical Unit	$\%$
Alternate Reported Units	kl in stressed areas
Normalization Conversion	Stressed volume / Total volume
Denominator Rule	Total Withdrawal
Core Flag	YES

`WAT-CON-001`

Canonical Metric Name	**Total Water Consumption**
XBRL Concept QName	`in-capmkt_TotalVolumeOfWaterConsumption`
Canonical Unit	m^3
Alternate Reported Units	kl
Normalization Conversion	$1\ \mathrm{kl} = 1\ \mathrm{m}^3$
Denominator Rule	Absolute (Withdrawal minus Discharge)
Core Flag	YES

`WAT-INT-001`

Canonical Metric Name	**Water Intensity per Turnover**
XBRL Concept QName	`in-capmkt_WaterIntensityPerRupeeOfTurnover`
Canonical Unit	$\mathrm{kl}/\mathrm{INR\ Cr}$
Alternate Reported Units	kl/INR
Normalization Conversion	Factor: $\times 10^{7}$
Denominator Rule	Turnover (INR Cr)
Core Flag	YES

`WAT-INT-002`

Canonical Metric Name	**Specific Water Consumption**
XBRL Concept QName	`in-capmkt_WaterIntensityInTermOfPhysicalOutput`
Canonical Unit	m^3/t
Alternate Reported Units	l/t, kl/t
Normalization Conversion	If l/t: $\div 1000$
Denominator Rule	Tonnes of Cementitious Product
Core Flag	YES

`WAT-DIS-001`

Canonical Metric Name	**Total Water Discharge**
XBRL Concept QName	`in-capmkt_TotalWaterDischargedInKilolitres`
Canonical Unit	m^3
Alternate Reported Units	kl
Normalization Conversion	$1\ \mathrm{kl} = 1\ \mathrm{m}^3$
Denominator Rule	Absolute (Cement plants typically target ZLD)
Core Flag	YES

`WAT-REC-001`

Canonical Metric Name	**Recycled / Reused Water Share**
XBRL Concept QName	`Corporate narrative / P6 tables`
Canonical Unit	%
Alternate Reported Units	%
Normalization Conversion	Recycled volume / Total consumption
Denominator Rule	Total Consumption
Core Flag	NO

Pillar III: EARTH & CIRCULARITY (Pillar Weight: 30%)

`EAR-CIR-001`

Canonical Metric Name	**Clinker Factor**
XBRL Concept QName	`Derived (Clinker produced / Cement produced)`
Canonical Unit	%
Alternate Reported Units	Pure ratio ($0.0 - 1.0$)
Normalization Conversion	If decimal: $\times 100$ (Industry avg ~65%-72%)
Denominator Rule	Total Cementitious Output
Core Flag	NO (Critical Benchmark)

`EAR-CIR-002`

Canonical Metric Name	**Alternative Raw Materials Used**
XBRL Concept QName	`Derived (Fly ash + Slag + Gypsum)`
Canonical Unit	Tonnes
Alternate Reported Units	Lakh MT, MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	NO (Critical Benchmark)

`EAR-WAS-001`

Canonical Metric Name	**Total Waste Generated**
XBRL Concept QName	`in-capmkt_TotalWasteGeneratedInMetricTonnes`
Canonical Unit	Tonnes
Alternate Reported Units	MT, kg
Normalization Conversion	If kg: $\div 1000$
Denominator Rule	Absolute
Core Flag	YES

`EAR-WAS-002`

Canonical Metric Name	**Hazardous Waste Generated**
XBRL Concept QName	`in-capmkt_TotalHazardousWasteGenerated`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	YES

`EAR-WAS-003`

Canonical Metric Name	**Plastic Waste Generated / Co-processed**
XBRL Concept QName	`in-capmkt_PlasticWaste`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	YES

`EAR-WAS-004`

Canonical Metric Name	**Waste Recovered / Recycled**
XBRL Concept QName	`in-capmkt_TotalWasteRecoveredThroughRecycling`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	YES

`EAR-WAS-005`

Canonical Metric Name	**Waste Disposed to Landfill**
XBRL Concept QName	`in-capmkt_WasteDisposedByLandfilling`
Canonical Unit	Tonnes
Alternate Reported Units	MT
Normalization Conversion	Standard metric tonnes
Denominator Rule	Absolute
Core Flag	YES

`EAR-AFR-001`

Canonical Metric Name	**Alternative Fuel Co-processed in Kilns**
XBRL Concept QName	`Corporate Sustainability / P6 Narrative`
Canonical Unit	Tonnes
Alternate Reported Units	Lakh MT, MT
Normalization Conversion	Standard metric tonnes (Municipal Solid Waste, RDF, Biomass)
Denominator Rule	Absolute
Core Flag	NO (Critical Benchmark)

`EAR-BIO-001`

Canonical Metric Name	**Quarry Rehabilitation / Green Cover**
XBRL Concept QName	`P6 Leadership Narrative / EIA reports`
Canonical Unit	Hectares
Alternate Reported Units	Acres, Hectares
Normalization Conversion	If Acres: $\times 0.404686$
Denominator Rule	Total Leased Mining Area
Core Flag	NO

---

3. Strict Comparability and Denominator Rules

1. **Clinker vs. Cement vs. Cementitious Material:**
   - *Clinker:* The kiln intermediate product before grinding and gypsum/fly-ash blending.
   - *Cementitious Material:* Clinker + Gypsum + Fly Ash (PPC) + Granulated Blast Furnace Slag (PSC) + limestone fillers.
   - **Rule:** Never compare $kgCO_2/\text{t clinker}$ against $kgCO_2/\text{t cementitious}$. Cementitious intensity is 20–35% lower due to clinker substitution. If the denominator is ambiguous, the observation is marked `WARNING: DENOMINATOR_AMBIGUOUS`.
2. **Consolidated vs. Standalone Boundary Shift:**
   - In FY2022–23, ACC and Ambuja filed BRSR on a 15-month *Consolidated* basis (following Adani Group acquisition and financial year alignment from Dec-year-end to March-year-end).
   - In FY2023–24, FY2024–25, and FY2025–26, both filed on a 12-month *Standalone* basis.
   - **Rule:** Flag any YoY comparison between FY23 and FY24 for ACC/Ambuja as `NOT_COMPARABLE: BOUNDARY_AND_PERIOD_SHIFT`.
3. **Double Counting Protection:**
   - Ambuja Cements owns ~50.05% of ACC Limited.
   - In consolidated sector benchmarks, ACC data must not be summed together with Ambuja consolidated figures without stripping the subsidiary overlap.

DOCUMENT 08

AWE Ingestion Architecture & Pipeline Specification

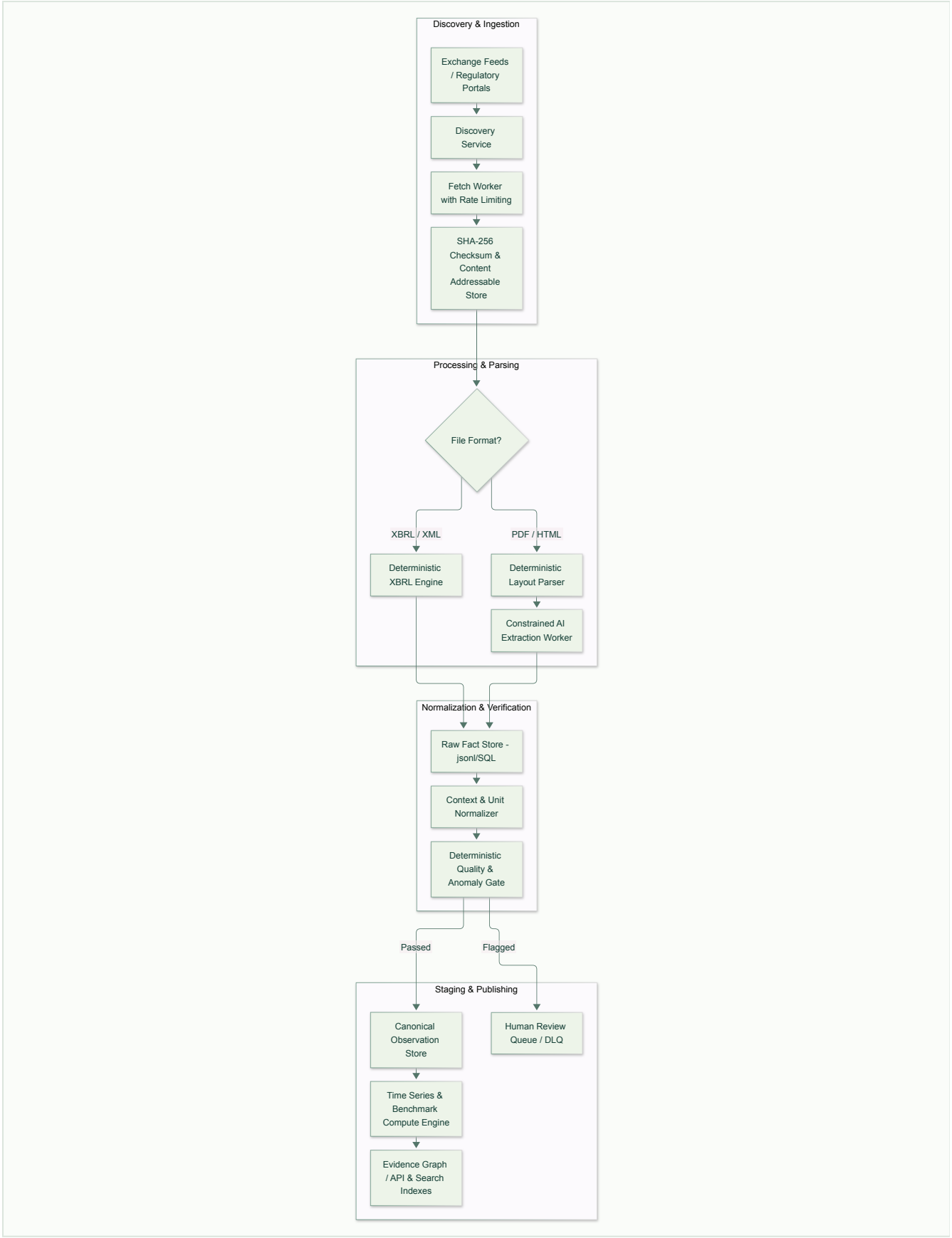
08-ingestion-architecture.md

Document ID: AWE-ARCH-INGEST-v1.0
Status: Canonical engineering architecture specification.
Version Anchor: AWE-PIPELINE-2026-v1.0

---

1. End-to-End Pipeline Flow

The ingestion pipeline is designed with a strict **provenance-first, append-only, idempotent** architecture. Every downstream data point must maintain a cryptographically auditable lineage back to its source byte payload.


Discovery & Ingestion
Exchange Feeds / Regulatory Portals
Discovery Service
Fetch Worker with Rate Limiting
SHA-256 Checksum & Content Addressable Store
Processing & Parsing
File Format?
XBRL / XML
PDF / HTML
Deterministic XBRL Engine
Deterministic Layout Parser
Constrained AI Extraction Worker
Normalization & Verification
Raw Fact Store - jsonl/SQL
Context & Unit Normalizer
Deterministic Quality & Anomaly Gate
Passed
Flagged
Staging & Publishing
Canonical Observation Store
Human Review Queue / DLQ
Time Series & Benchmark Compute Engine
Evidence Graph / API & Search Indexes

2. Ingestion Pipeline Stages

Stage 1: DISCOVER & MANIFEST

- Polls official repositories (NSE Corporate Filings, BSE Listing Center) using scheduled workers.
- Generates a discovery manifest record containing: source exchange, filing ID, submission timestamp, security symbol, and download URL.
- Records all HTTP headers, redirect chains, and response codes.

Stage 2: DOWNLOAD & HASH (Immutable Raw Storage)

- Downloads the raw binary payload to content-addressable storage: `storage/raw/sha256/{hash[:2]}/{hash[2:4]}/{hash}.{ext}`
- Computes SHA-256 hash immediately upon byte stream completion.
- Verifies byte size against `Content-Length`.
- **Write-Once Invariant:** If an identical SHA-256 hash already exists in storage, the binary is not overwritten; instead, a new `retrieval` event is appended referencing the existing `raw_object`.

Stage 3: DEDUPLICATE & IDENTIFY

- Identifies corporate identity from embedded XML tags (`CorporateIdentityNumber`, `ISIN`, `LegalName`, `NSESymbol`) rather than trusting URL slugs or filenames.
- Resolves reporting period dates (`startDate`, `endDate`, `instant`) and context bounds.
- Checks `report_version` table: if filing is marked as an amendment, flags the superseded filing ID and retains both versions with a `supersedes_id` relationship.

Stage 4: PARSE (XBRL First, AI Second)

- **XBRL Engine:**
  - Evaluates root `schemaRef` to pin the exact SEBI taxonomy version (2021, 2023, 2024, 2025, or 2026).
  - Uses streaming XML parsers (`lxml.etree.iterparse`) with disabled external entity resolution (`resolve_entities=False`, `network=False`) to prevent XXE vulnerabilities.
  - Extracts 100% of facts, associating each with its resolved context (dates, explicit dimensions, typed dimensions) and unitRef (measures, numerators, denominators).
  - Preserves exact XPath location (`/*/*[ordinal]`) and lexical string value.
- **Constrained AI Extraction (PDFs and Narratives):**
  - Triggered only for PDF-only filings, qualitative target statements, assurance scopes, and contextual footnotes.
  - Strict JSON Schema output enforcement:

```
{
  "extracted_value": "number or null",
  "original_text": "string verbatim quote",
  "source_page": 142,
```

```
  "table_index": 2,
  "bounding_box": [ymin, xmin, ymax, xmax],
  "extraction_confidence": 0.95,
  "reasoning_category": "TARGET_COMMITMENT",
  "needs_human_review": false
}
```

- **Inviolable AI Rule:** If a number is absent in the text, `extracted_value` must be `null`. The LLM is strictly prohibited from estimating, extrapolating, or synthesizing values.

Stage 5: NORMALIZE & HARMONIZE

- Applies versioned conversion formulas:
  - Water: $\text{kl} \to \text{m}^3$ $(1:1)$, $\text{l} \to \text{m}^3$ $(\div 1000)$.
  - Energy: $\text{MWh} \to \text{GJ}$ $(\times 3.6)$, $\text{TJ} \to \text{GJ}$ $(\times 1000)$.
  - Mass: $\text{kg} \to \text{t}$ $(\div 1000)$, $\text{Lakh MT} \to \text{t}$ $(\times 100,000)$.
  - Scale: Handles Million / Lakh scales based on explicit XBRL units or documented entity scale.
- Retains original value, original unit, conversion factor, normalized value, and normalized unit.

Stage 6: VALIDATE & ANOMALY DETECTION

- Runs deterministic rules against normalized facts (e.g. Scope balance, water balance, boundary shifts).
- Assigns quality status: `VALID`, `WARNING`, `ERROR`, or `MANUAL_REVIEW`.

Stage 7: VERSION & PUBLISH

- Commits normalized observations to the queryable database with foreign keys to `calculation_run` and `methodology_version`.
- Updates materialized views, peer percentiles, and full-text/vector search indices.

---

3. Reliability, Idempotency, and Failure Queues

1. **Idempotency Guarantee:** The composite key `(document_sha256, source_location, parser_version)` enforces uniqueness. Running the ingestion worker N times against the same file yields exactly the same fact rows with zero duplicates.
2. **Dead Letter Queue (DLQ):** When XML validation fails, an unknown namespace is encountered, or JSON schema validation fails in AI extraction:
   - The job is routed to `ingestion_dlq` table with full stack trace, raw payload snapshot, attempt counter, and error categorization.
   - Alerts are dispatched to the engineering queue.
3. **Exponential Backoff:** Transitory network errors during retrieval use exponential backoff:

$$t_{wait} = \min\left(t_{max}, t_{base} \cdot 2^{attempt}\right) \pm \text{jitter}$$

   Max attempts = 5. After 5 failures, the job enters DLQ for operator inspection.

DOCUMENT 09

AWE Canonical Database Schema Specification

09-database-schema.md

Document ID: AWE-DB-SPEC-v1.0
Companion File: 09-database-schema.sql
Target Engine: PostgreSQL 15+ / TimescaleDB / AWS Aurora PostgreSQL

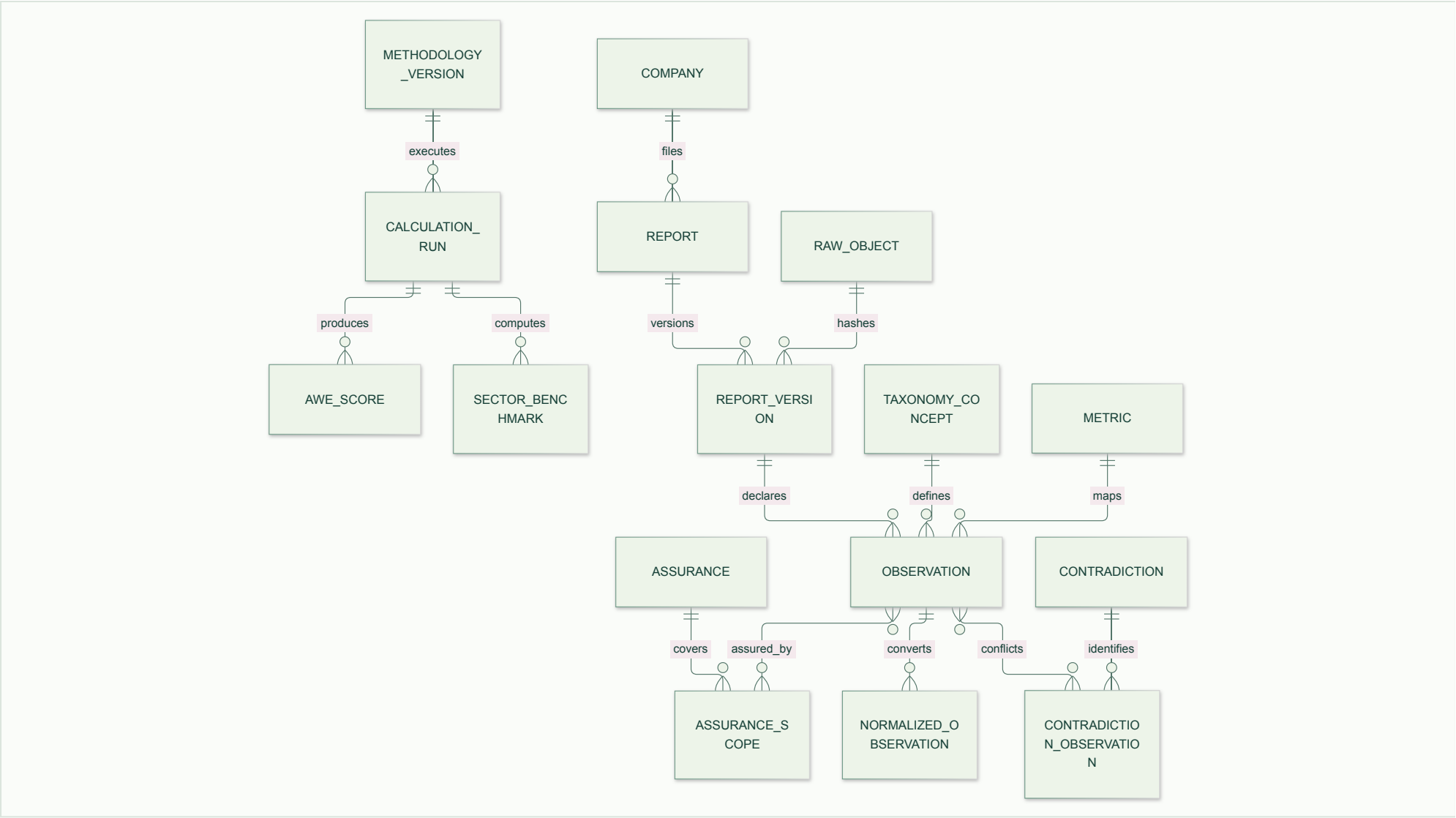
---

1. Schema Design Principles

1. **Multi-Observation Disagreement Model:** Traditional enterprise schemas enforce `UNIQUE(company_id, metric_id, year)` to store a single "truth" value. AWE strictly rejects this model. If a company reports water consumption as 4.2 million m³ in its FY24 BRSR, but reports the FY24 comparative as 3.8 million m³ in its FY25 filing, **both assertions are preserved**. They are linked via `contradiction` and `contradiction_observation` tables with resolution notes.
2. **Strict Cryptographic Lineage:** Every single observation in `observation` joins directly to `report_version`, which joins to `raw_object` via `document_sha256`. An observation cannot exist in the system without an immutable raw document hash and a precise `source_location` (XPath or page index).
3. **Separation of Extracted vs. Normalized Data:** `observation` stores the raw string (`original_lexical_value`) and parsed number (`original_numeric_value`) alongside raw unit strings. `normalized_observation` contains the converted value in canonical SI units, recording the exact `conversion_factor_applied`.

---

2. Entity Groupings


METHODOLOGY _VERSION
COMPANY
executes
files
CALCULATION_ RUN
REPORT
RAW_OBJECT
produces
computes
versions
hashes
AWE_SCORE
SECTOR_BENC HMARK
REPORT_VERSI ON
TAXONOMY_CO NCEPT
METRIC
declares
defines
maps
ASSURANCE
OBSERVATION
CONTRADICTION
covers
assured_by
converts
conflicts
identifies
ASSURANCE_S COPE
NORMALIZED_O BSERVATION
CONTRADICTIO N_OBSERVATIO N

Core Entity Tables

- **Corporate:** `company`, `company_identifier` (CIN, LEI), `company_alias`, `security` (ISIN), `listing` (NSE, BSE), `sector`, `company_relationship` (subsidiary/holding), `facility`.
- **Evidence & Storage:** `rights_policy`, `source`, `raw_object` (content-addressable by SHA-256), `retrieval` (HTTP logs).
- **XBRL & Reporting:** `reporting_period`, `reporting_boundary`, `taxonomy_version`, `taxonomy_concept`, `report`, `report_version`, `xbrl_context`, `xbrl_unit`.
- **Metrics & Observations:** `unit`, `metric`, `concept_metric_mapping`, `observation`, `normalized_observation`.
- **Integrity & Intelligence:** `assurance`, `assurance_scope`, `target`, `contradiction`, `contradiction_observation`, `quality_issue`.
- **Scoring & Benchmarks:** `methodology_version`, `calculation_run`, `sector_benchmark`, `awe_score`.

DOCUMENT 10

AWE Automated Data Quality Rules & Validation Engine

10-data-quality-rules.md

Document ID: AWE-QUAL-ENG-v1.0
Status: Canonical engineering rule catalog.
Applicability: Ingestion pipeline, continuous data auditing, and anomaly detection.

---

1. Severity Tiers and Action Protocols

Tier	Meaning	Action Protocol
`VALID`	Observation satisfies all physical bounds, accounting identities, and taxonomy rules.	Committed directly to canonical observation store and eligible for benchmark computation.
`WARNING`	Minor mathematical discrepancy, missing footnote, or non-material divergence (<5%).	Committed to canonical store with warning badge; included in benchmarks with sensitivity flag.
`MANUAL_REVIEW`	Plausible anomaly: >50% YoY jump, boundary change (e.g. Consolidated to Standalone), or restatement.	Queued for expert analyst review. Excluded from automated score publication until reviewed.
`ERROR`	Severe corruption: negative mass/energy, percentage > 100%, unresolvable units, broken XML tags.	Quarantined. Raw fact retained for audit, but excluded completely from calculations.

> [!CRITICAL] **No Silent Corrections:** The system will **never** silently recalculate or overwrite an issuer's reported value. If a company mistakenly reports water withdrawal in liters instead of kiloliters, the raw assertion remains untouched, and a normalized assertion is created accompanied by an explicit `quality_issue` record.

---

2. Rule Catalog

Group A: Corporate Identity & Filing Integrity

`ID-001`

Rule Name	**CIN Format Validation**
Expression / Condition	`CIN ~ '^[L
Severity	U][0-9]{5}[A-Z]{2}[0-9]{4}[A-Z]{3}[0-9]{6}$'`
Description	`ERROR`

`ID-002`

Rule Name	**ISIN Structure Validation**
Expression / Condition	`ISIN ~ '^IN[A-Z0-9]{9}[0-9]$'`
Severity	`ERROR`
Description	Verifies Indian 12-character alphanumeric security identifier with MOD-10 check digit.

`FIL-001`

Rule Name	**Period Continuity Check**
Expression / Condition	`DurationDays = (endDate - startDate) + 1`
Severity	`WARNING`
Description	Verifies duration. If duration is not 365/366 days (e.g. 455 days for ACC/Ambuja FY23), flags period anomaly.

`FIL-002`

Rule Name	**Duplicate Filing Check**
Expression / Condition	`COUNT(*) OVER (PARTITION BY CIN, period_id, report_type) > 1`
Severity	`WARNING`
Description	Identifies duplicate or amended submissions for the same period.

Group B: Physical Bounds & Thermodynamic Realities

`PHY-001`

Rule Name	**Non-Negative Energy**
Expression / Condition	`TotalEnergyConsumptionInJoules >= 0`
Severity	`ERROR`
Description	Total energy consumption cannot be negative under thermodynamic laws.

`PHY-002`

Rule Name	**Non-Negative Mass**
Expression / Condition	`Scope1Emissions >= 0 AND Scope2Emissions >= 0`
Severity	`ERROR`
Description	Gross emissions cannot be negative. Removals/offsets must be disclosed under separate negative accounts.

`PHY-003`

Rule Name	**Percentage Range Bound**
Expression / Condition	`0.0 <= PercentageValue <= 100.0`
Severity	`ERROR`
Description	Concepts typed as percentages (e.g. renewable share, recycled waste share) cannot exceed 100%.

`PHY-004`

Rule Name	**Water Balance Identity**
Expression / Condition	`TotalWithdrawal >= TotalConsumption`
Severity	`WARNING`
Description	Water withdrawal must equal or exceed consumption under standard hydrological mass balance (unless significant prior storage drawdown is explicitly footnoted).

`PHY-005`

Rule Name	**Waste Balance Check**
Expression / Condition	`TotalGenerated >= (TotalRecovered + TotalDisposed)`
Severity	`WARNING`
Description	Total waste generated should reconcile with the sum of recovered and disposed fractions within a 5% tolerance.

Group C: Accounting & Reconciliation Identities

`ACC-001`

Rule Name	**Scope 1+2 Additivity**
Expression / Condition	`ABS((Scope1 + Scope2) - TotalScope1and2) / TotalScope1and2 <= 0.01`
Severity	`WARNING`
Description	Combined emissions must reconcile with individual Scope sums within 1% rounding error.

`ACC-002`

Rule Name	**Water Source Additivity**
Expression / Condition	`ABS(SUM(Groundwater + Surface + ThirdParty + Seawater + Others) - TotalWithdrawal) / TotalWithdrawal <= 0.01`
Severity	`WARNING`
Description	Total water withdrawal must reconcile with disaggregated source volumes.

`ACC-003`

Rule Name	**Waste Stream Additivity**
Expression / Condition	`ABS(SUM(Hazardous + NonHazardous) - TotalWasteGenerated) / TotalWasteGenerated <= 0.02`
Severity	`WARNING`
Description	Total waste must reconcile with hazardous and non-hazardous breakdowns.

Group D: Longitudinal Anomaly & Restatement Detection

`LONG-001`

Rule Name	**Sudden YoY Intensity Shift**
Expression / Condition	`ABS(Intensity_t - Intensity_{t-1}) / Intensity_{t-1} > 0.50`
Severity	`MANUAL_REVIEW`
Description	Flags any single-year shift greater than 50% in GHG or water intensity. Triggers narrative search for M&A or plant closure.

`LONG-002`

Rule Name	**Historical Comparative Restatement**
Expression / Condition	`ReportedPriorValue_{t} != OriginalReportedValue_{t-1}`
Severity	`MANUAL_REVIEW`
Description	Identifies historical restatements between filings. Automatically creates a `contradiction` record.

`LONG-003`

Rule Name	**Boundary Shift Detection**
Expression / Condition	`BoundaryBasis_t != BoundaryBasis_{t-1}`
Severity	`MANUAL_REVIEW`
Description	Flags transition between Standalone and Consolidated reporting (e.g. ACC FY23 Consolidated → FY24 Standalone).

`LONG-004`

Rule Name	**Repeated Identical Values**
Expression / Condition	`Value_t == Value_{t-1} AND Value_{t-1} == Value_{t-2}`
Severity	`WARNING`
Description	Detects boilerplate copy-pasting of identical numbers across multiple consecutive financial years.

DOCUMENT 11

AWE Public & Research API Specification (v1.0)

11-api-specification.md

Document ID: AWE-API-SPEC-v1.0
Specification Format: OpenAPI 3.1 compatible
Base URL: `https://api.aweindia.org/v1`
Authentication: Read-only public access for published metrics; API Key for bulk research exports.

---

1. Architectural Philosophy

Every API response returning a quantitative value must include:

1. Canonical normalized value and unit.
2. Original reported value and raw unit string.
3. Cryptographic document hash (`document_sha256`) and source locator.
4. Comparability flag (`HIGH`, `MEDIUM`, `LOW`, `NOT_COMPARABLE`).
5. TRS Evidence Confidence tier (`E0` to `E5`).

---

2. Core Endpoints

2.1 Corporate Directory & Profiles

`GET /companies`

Returns paginated list of tracked corporate entities.

- **Parameters:**
  - `sector` (string, optional): e.g. `cement`
  - `min_coverage` (float, optional): e.g. `0.75`
  - `limit` (int, default: 20, max: 100)
  - `offset` (int, default: 0)
- **Response (200 OK):**

```
{
  "total": 6,
  "limit": 20,
  "offset": 0,
  "data": [
    {
      "company_id": "9b1deb4d-3b7d-4bad-9bdd-2b0d7b3dcb6d",
      "legal_name": "J.K. Cement Limited",
```

```
      "cin": "L17229UP1994PLC017199",
      "isin": "INE823G01014",
      "nse_symbol": "JKCEMENT",
      "bse_code": "532644",
      "sector": "Cement",
      "latest_score": {
        "period": "FY2025-26",
        "awe_performance": 74.2,
        "evidence_confidence": 92.5,
        "score_coverage": 87.5
      }
    }
  ]
}
```


`GET /companies/{company_id}`

Returns complete profile, pillars, and governance metadata for a company.

2.2 Historical Time Series & Comparability

`GET /companies/{company_id}/time-series`

Returns multi-year environmental trajectory for a specific metric.

- **Parameters:**
  - `metric_code` (string, required): e.g. `AIR-GHG-001` (Scope 1 Emissions)
  - `start_fy` (int, optional): e.g. `2022`
  - `end_fy` (int, optional): e.g. `2026`
- **Response (200 OK):**


```
{
  "company_id": "9b1deb4d-3b7d-4bad-9bdd-2b0d7b3dcb6d",
  "company_name": "JK Cement Limited",
  "metric": {
    "code": "AIR-GHG-001",
    "name": "Gross Scope 1 GHG Emissions",
    "canonical_unit": "tCO2e"
  },
  "time_series": [
    {
      "financial_year": "FY2022-23",
      "start_date": "2022-04-01",
      "end_date": "2023-03-31",
      "normalized_value": 7985999.42,
      "normalized_unit": "tCO2e",
      "original_reported_value": "7985999.42",
      "original_reported_unit": "Non-SI:tCO2e",
      "reporting_boundary": "STANDALONE",
      "comparability": "HIGH",
      "evidence_confidence_tier": "E4",
      "provenance": {
        "document_sha256": "4090fa71c05ec1b1c7e9563f2ce38b2faaa09a0dcc7a2f67ae0a3f875d1acd2a",
        "source_url":
"https://nsearchives.nseindia.com/corporate/xbrl/BRSR_933208_22092023043743_WEB.xml",
```

```
        "xpath": "/*/*[602]"
      }
    },
    {
      "financial_year": "FY2023-24",
      "start_date": "2023-04-01",
      "end_date": "2024-03-31",
      "normalized_value": 10334366.0,
      "normalized_unit": "tCO2e",
      "original_reported_value": "10334366",
      "original_reported_unit": "Non-SI:tCO2e",
      "reporting_boundary": "STANDALONE",
      "comparability": "HIGH",
      "evidence_confidence_tier": "E5",
      "assurance_provider": "DNV Business Assurance India Private Limited"
    },
    {
      "financial_year": "FY2024-25",
      "start_date": "2024-04-01",
      "end_date": "2025-03-31",
      "normalized_value": 10472587.0,
      "normalized_unit": "tCO2e",
      "original_reported_value": "10472587",
      "original_reported_unit": "Non-SI:tCO2e",
      "reporting_boundary": "STANDALONE",
      "comparability": "HIGH",
      "evidence_confidence_tier": "E5",
      "assurance_provider": "TUV India Private Limited"
    },
    {
      "financial_year": "FY2025-26",
      "start_date": "2025-04-01",
      "end_date": "2026-03-31",
      "normalized_value": 12725657.0,
      "normalized_unit": "tCO2e",
      "original_reported_value": "12725657",
      "original_reported_unit": "Non-SI:tCO2e",
      "reporting_boundary": "STANDALONE",
      "comparability": "HIGH",
      "evidence_confidence_tier": "E5",
      "assurance_provider": "TUV India Private Limited"
    }
  ]
}
```

2.3 Target & Commitment Ledger

`GET /companies/{company_id}/targets`

Returns verified environmental commitments and formulaic progress tracking.

- **Response (200 OK):**


```
{
  "company_id": "9b1deb4d-3b7d-4bad-9bdd-2b0d7b3dcb6d",
  "targets": [
    {
```

```
      "target_id": "550e8400-e29b-41d4-a716-446655440000",
      "metric_code": "AIR-INT-002",
      "metric_name": "Specific Net CO2 per Tonne Cementitious Material",
      "baseline_year": 2020,
      "baseline_value": 612.0,
      "target_year": 2030,
      "target_value": 465.0,
      "current_year": 2026,
      "current_value": 521.0,
      "target_unit": "kgCO2e/t",
      "progress_percentage": 61.9,
      "expected_progress_percentage": 60.0,
      "trajectory_status": "ON_TRAJECTORY",
      "sbti_status": "1.5C_VALIDATED"
    }
  ]
}
```

2.4 Sector Benchmarks & Distributions

`GET /sectors/{sector_id}/benchmarks`

Returns robust statistical distributions across sector cohorts.

- **Parameters:**
  - `metric_code` (string, required): e.g. `AIR-INT-002`
  - `period_label` (string, required): e.g. `FY2025-26`
- **Response (200 OK):**


```
{
  "sector": "Cement",
  "metric_code": "AIR-INT-002",
  "period": "FY2025-26",
  "unit": "kgCO2e/t",
  "sample_n": 6,
  "statistics": {
    "p10": 498.2,
    "p25": 512.4,
    "median_p50": 538.0,
    "p75": 564.1,
    "p90": 589.6,
    "mean": 540.3,
    "std_dev": 32.1
  },
  "methodology_version": "AWE-CEMENT-2026-v1.0"
}
```

2.5 Discrepancies & Contradictions

`GET /companies/{company_id}/contradictions`

Returns flagged discrepancies, historical restatements, and boundary shifts.

- **Response (200 OK):**

```
{
  "company_id": "acc-cement-uuid",
  "contradictions": [
    {
      "contradiction_id": "c-001",
      "classification": "PERIOD_DIFFERENCE",
      "materiality": "HIGH",
      "metric_code": "AIR-GHG-001",
      "description": "FY2022-23 reported across 15 months (2022-01-01 to 2023-03-31) on
Consolidated basis following Adani takeover; FY2023-24 reported across 12 months on Standalone
basis.",
      "resolution_status": "FLAGGED_INCOMPARABLE"
    }
  ]
}
```

2.6 Natural Language & Structured Query Interface

`POST /query`

Receives natural-language questions, compiles them into safe, parameterized SQL, and executes deterministically.

- **Request Body:**

```
{
  "query": "Which cement companies reduced carbon intensity every year for the last 3 years?"
}
```

- **Response (200 OK):**

```
{
  "compiled_sql": "SELECT c.legal_name, m.canonical_code, ... WHERE ...",
  "execution_time_ms": 14,
  "results": [
    {
      "company": "Dalmia Bharat Limited",
      "intensity_fy23": 508.0,
      "intensity_fy24": 495.0,
      "intensity_fy25": 489.0,
      "intensity_fy26": 481.0,
      "trend": "STRICT_ANNUAL_DECREASE"
    }
  ],
  "explanation": "Dalmia Bharat Limited is the only company among the six studied that achieved
an uninterrupted annual reduction in specific net carbon intensity per tonne from FY23 to FY26."
}
```

DOCUMENT 12

AWE Dual-Index Search & Evidence-Grounded RAG Architecture

12-search-rag-architecture.md

Document ID: AWE-ARCH-RAG-v1.0
Status: Engineering architecture specification.
Objective: Provide hallucination-free natural language research capabilities grounded in deterministic data and verifiable text chunks.

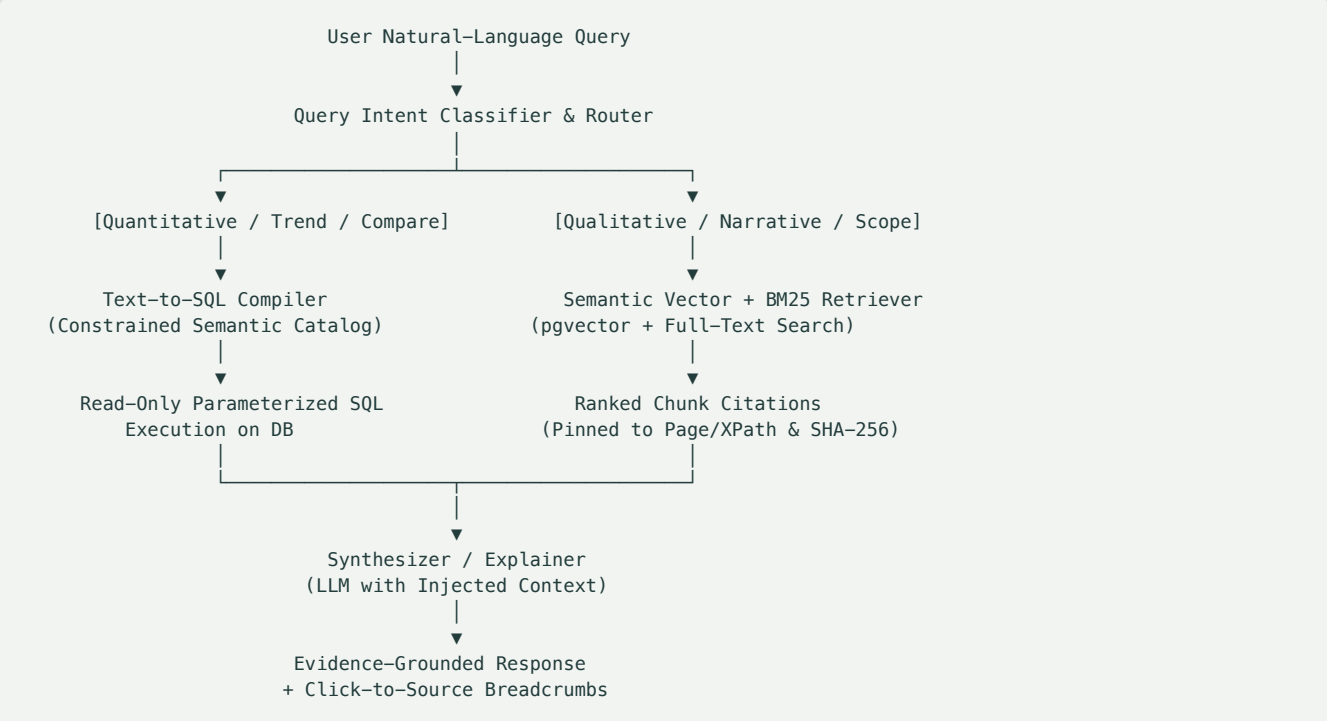
---

1. Dual-Index Philosophy

A fundamental flaw of conventional ESG search interfaces is passing unstructured text chunks into a general-purpose LLM to ask "What was Company X's emissions in 2024?". This frequently yields hallucinations, missed restatements, and erroneous unit conversions.

AWE divides all environmental intelligence into two strict indices:

1. **The Deterministic Structured Metric Index (PostgreSQL/TimescaleDB):** All quantitative facts, time series, financial denominators, targets, and benchmarks.
2. **The Cryptographic Evidence Vector & Text Index (PostgreSQL pgvector / BM25):** All qualitative disclosures, assurance opinions, methodological appendices, target narratives, and facility descriptions.



---

2. Text-to-SQL Semantic Layer

To ensure mathematical precision and eliminate SQL injection:

- The user's query is never converted into arbitrary SQL.
- Queries are compiled against a **whitelisted semantic schema**:
  - `v_company_time_series(company_name, metric_code, fy, value, unit, boundary, comparability)`
  - `v_target_progress(company_name, metric_name, baseline_fy, target_fy, progress_pct, status)`
  - `v_sector_benchmarks(sector, metric_name, fy, p10, p25, median, p75, p90)`
- The SQL generator is constrained by:
  - Strict read-only transaction mode (`SET TRANSACTION READ ONLY;`).
  - Strict timeout limit (`SET statement_timeout = '2000ms';`).
  - Mandatory `LIMIT 100` clause.
- If the query cannot be expressed via the semantic view catalog, the engine rejects the request with a clear explanation rather than fabricating arbitrary SQL.

---

3. Evidence Vector Chunking Specification

Every page of regulatory PDF text and every Section C narrative block from XBRL filings is chunked and indexed:

3.1 Chunk Data Schema

```
{
  "chunk_id": "a98f12c3-4d5e-6f7a-8b9c-0d1e2f3a4b5c",
  "document_sha256": "ca0fd08855faf611ad2bd68f8937ba935af8aff0d9ef93fad818f475c27ff170",
  "company_id": "ultratech-cement-uuid",
  "financial_year": "FY2025-26",
  "source_url":
"https://nsearchives.nseindia.com/corporate/xbrl/BRSR_XML_BRSR_WebXMLFile_20260724_121711945.xml",
  "page_number": 84,
  "section_reference": "Principle 6, Essential Indicator 7: Scope 1 GHG emissions",
  "chunk_text": "Scope 1 emissions include direct emissions from fuel combustion in cement rotary
kilns, captive thermal power plants, and calcination of limestone. Process emissions constitute 62%
of direct emissions...",
  "embedding": [0.0124, -0.0452, ..., 0.0891],
  "verified_assurance_provider": "TUV SUD South Asia Private Limited",
  "chunk_hash": "e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855"
}
```

3.2 Hybrid Retrieval Strategy

- Retrieval employs **Reciprocal Rank Fusion (RRF)** combining:
  1. Dense vector similarity (cosine distance on `pgvector` HNSW index).
  2. Sparse keyword matching (`ts_rank_cd` over PostgreSQL English dictionary).

- Hybrid search guarantees retrieval of exact regulatory acronyms (e.g. `AFR`, `WHRS`, `EPR`, `PAT Scheme`, `TSR`) which pure embedding models frequently dilute.

---

4. Citation and Verification Contract

Every LLM-generated narrative response must include interactive citation tags:

> Dalmia Bharat achieved a Thermal Substitution Rate (TSR) of 18.2% in FY2025-26 [^1], primarily driven by biomass and municipal RDF co-processing at its Belgaum and Dalmiapuram kilns [^2].
>
> [^1]: Dalmia Bharat FY26 BRSR, Section C, P6 E1 [^2]: TUV India Assurance Report, Page 4, Paragraph 2

Clicking the citation immediately displays the exact raw XML snippet or highlight on the verified PDF page.

DOCUMENT 13

AWE Public Visualization Library Specification

13-visualization-library.md

Document ID: AWE-VIZ-LIB-v1.0
Status: Public visual journalism & research visualization catalog.
Version Anchor: AWE-VIZ-2026-v1.0

---

1. Visualization Integrity Principles

Every chart rendered in AWE must adhere to five non-negotiable data integrity rules:

1. **Source & SHA-256 Attribution:** Every visualization card displays the primary regulatory authority, filing IDs, and methodology version in its footer.
2. **Boundary Discontinuity Handling:** When a company undergoes a period shift (e.g. 15-month FY23 for ACC/Ambuja) or reporting boundary change (Consolidated $\leftrightarrow$ Standalone), the trend line **must visually break** with a dashed/dotted connector and a prominent warning tooltip.
3. **No Dual-Axis Distortion:** Incompatible units (e.g. absolute emissions and revenue) must not be plotted on misleading overlaid dual-Y axes with arbitrary scaling.
4. **Outlier & Sample Caveats:** Sector median lines must state the exact sample size ($n = 6$) and exclude incomplete entities.
5. **Traceability:** Hovering or clicking on any point/bar opens a modal displaying the exact raw reported fact, conversion factor, and XBRL XPath.

---

2. The 10 Core Visualizations

Viz 01: Absolute Scope 1 Decarbonisation Trajectory (FY22–FY26)

- **Chart Type:** Multi-series Line Chart with Discontinuity Markers.
- **Metric:** `AIR-GHG-001` (Gross Scope 1 GHG Emissions).
- **Unit:** $\text{Million tCO}_2\text{e}$ (MtCO_2e).
- **Entities:** UltraTech Cement, Ambuja Cements, ACC, Shree Cement, Dalmia Bharat, JK Cement.
- **Comparability Rule:** ACC and Ambuja FY23 points are rendered with diamond markers and dashed connectors denoting the 15-month transition period. Dalmia Bharat FY22 value is normalized from reported raw tonnes ($12.55\ \text{Mt}$).
- **Key Insight:** UltraTech's emissions grew from 56.6 Mt (FY21) to 88.9 Mt (FY26) driven by massive capacity additions, highlighting the divergence between absolute growth and specific intensity reduction.

Viz 02: Scope 2 Grid Decarbonisation & Renewable Power Shift

- **Chart Type:** Stacked Area Chart / 100% Stacked Bar Chart.
- **Metric:** Grid Electricity vs. Captive Renewable Electricity vs. Waste Heat Recovery (WHRS).
- **Unit:** $\%$ of Total Electrical Consumption.
- **Key Insight:** Accelerated deployment of WHRS and solar/wind PPAs across Shree Cement and JK Cement, expanding renewable electricity shares above 40%.

Viz 03: Carbon Intensity: Physical vs. Financial Denominators

- **Chart Type:** Dual Small-Multiples Scatter/Line Comparison.
- **Series A:** Specific Net CO2 Intensity per Tonne of Cementitious Output ($kgCO_2e/t$).
- **Series B:** GHG Intensity per Rupee of Turnover ($tCO_2e/\text{INR Cr}$).
- **Key Insight:** Revenue intensity fluctuates wildly with cement market price volatility and inflation, whereas physical intensity ($kgCO_2/t$) accurately reflects kiln thermal efficiency and clinker factor reductions.

Viz 04: Industrial Thermal Energy Mix (Fossil vs. Alternative Fuels)

- **Chart Type:** Stacked Bar Chart.
- **Metric:** Kiln Thermal Fuel Breakdown (Petcoke, Imported Coal, Domestic Coal, Biomass, Municipal RDF).
- **Unit:** Terajoules (TJ) and Thermal Substitution Rate (TSR %).
- **Key Insight:** Dalmia Bharat and UltraTech scaling TSR toward 15–20% via aggressive co-processing of refuse-derived fuel (RDF) and industrial biomass.

Viz 05: Specific Water Consumption & Stress Overlap Matrix

- **Chart Type:** Grouped Bar Chart with Water-Stress Heatmap Overlay.
- **Metric:** Specific Water Consumption ($m^3/\text{tonne cement}$) alongside % Withdrawal from CGWB/WRI Water-Stressed Basins.
- **Unit:** m^3/t and $\%$ Stressed.
- **Key Insight:** Cement plants located in arid regions (Rajasthan, Gujarat) exhibit high water-stressed extraction (>60%), making rainwater harvesting and dry-process air-cooled condensers critical.

Viz 06: Water Withdrawal by Hydrological Source

- **Chart Type:** Treemap / Sankey Diagram.
- **Metric:** Groundwater vs. Surface Water vs. Third-Party Municipal vs. Harvested Rainwater.
- **Unit:** Million Cubic Meters (Mm^3).

- **Key Insight:** Heavy reliance on groundwater extraction across Northern and Western cluster plants, with gradual transition toward harvested quarry pit water.

Viz 07: Circular Economy: Waste Generation vs. Co-Processing (AFR)

- **Chart Type:** Bi-Directional (Tornado) Bar Chart.
- **Left Wing:** Waste Generated Internally (Hazardous + Non-hazardous + Refractory waste).
- **Right Wing:** External Industrial Waste Consumed / Co-processed (Fly ash, Slag, Plastic waste).
- **Key Insight:** Indian cement industry is a net waste-consumer ("circular sink"), consuming over $5\times$ more fly-ash and blast-furnace slag than the total solid waste it generates.

Viz 08: Corporate Environmental Target Ledger: Commitments vs. Trajectory

- **Chart Type:** Multi-Point Target Progress Dumbbell / Bullet Chart.
- **Elements per Company:**
  - Baseline Year & Value (e.g. 2020: $612\ \mathrm{kgCO_2/t}$)
  - Current FY26 Value (e.g. $521\ \mathrm{kgCO_2/t}$)
  - 2030 Interim Target (e.g. $465\ \mathrm{kgCO_2/t}$)
  - 2050 Net-Zero Long-term Target
- **Trajectory Status:** Color-coded badges (`ON TRAJECTORY`, `NEAR TRAJECTORY`, `OFF TRAJECTORY`).

Viz 09: The AWE Quadrant: Environmental Performance vs. Evidence Confidence

- **Chart Type:** 2D Scatter Quadrant Chart.
- **X-Axis:** AWE Performance Score ($0 - 100$).
- **Y-Axis:** TRS Evidence Confidence Score ($0 - 100\%$).
- **Quadrants:**
  - *High Performance + High Evidence:* **Verified Environmental Leader**
  - *Low Performance + High Evidence:* **Audited High-Impact Operator** (Clean data, dirty operations)
  - *High Performance + Low Evidence:* **Unverified Green Claims** (High greenwashing risk)
  - *Low Performance + Low Evidence:* **Critical Risk Entity**

Viz 10: Five-Year Evolution of Statutory Disclosure & Assurance (FY22–FY26)

- **Chart Type:** Multi-Layered Heatmap / Gantt Matrix.
- **Dimensions:** Company $\times$ Financial Year.
- **Cells Display:**
  - Filing Format: Voluntary PDF $\rightarrow$ Mandatory XBRL $\rightarrow$ BRSR Core.
  - Assurance Level: None $\rightarrow$ Management Assessment $\rightarrow$ Limited Assurance $\rightarrow$ Reasonable Assurance.

    - Assurance Provider: TUV India, SGS, BDO, DNV.
- **Key Insight:** Demonstrates the dramatic maturation of Indian regulatory disclosures from FY22 (voluntary, zero assurance) to FY26 (statutory reasonable assurance under SEBI Core mandate).

DOCUMENT 14

AWE Company Profile Specification & UI Wireframe

14-company-profile-wireframe.md

Document ID: AWE-UI-COMP-v1.0
Route: `/company/{slug}` (e.g. `/company/jk-cement`)
Design Paradigm: Evidence-first, auditable, high-density scientific dashboard.

---

1. Top Hero Section: The Environmental Triad

```
┌──────────────────────────────────────────────────────────────────────────────┐
│  J.K. Cement Limited                                  CIN: L17229UP1994PLC017199          │
│  Sector: Cement & Clinker (NIC 23941)                 ISIN: INE823G01014 | NSE: JKCEMENT  │
│  Reporting FY: 2025-26 (Standalone Basis)             Filing: NSE XBRL (Verified Primary) │
├──────────────────────┬──────────────────────┬────────────────────────────────┤
│      AWE PERFORMANCE SCORE      │     EVIDENCE CONFIDENCE     │             SCORE COVERAGE             │
│              74.2               │            94.0%            │               88.5%                    │
│        [ Sector Rank: 2/6 ]     │     Tier: E5 (Reasonable)   │     23 of 26 Core Metrics Valid        │
├──────────────────────┴──────────────────────┴────────────────────────────────┤
│  Pillar Breakdown:                                                           │
│  ├── [AIR]   68.5/100 (Weight: 45%)  ── Net Scope 1: 542 kgCO2/t | Renewable Elec: 44.1%          │
│  ├── [WATER] 91.0/100 (Weight: 25%)  ── Specific Water: 0.142 m3/t | Zero Liquid Discharge Target Met  │
│  └── [EARTH] 78.4/100 (Weight: 30%)  ── TSR: 14.8% | Clinker Factor: 65.2% | Waste Recycled: 96.2%    │
└──────────────────────────────────────────────────────────────────────────────┘
```

---

2. Interactive Analytical Sections

Tab 1: Longitudinal Trends & Comparability

- Displays 5-Year historical trends (FY22 to FY26) for:
  - Gross & Net Scope 1 GHG Emissions ($\mathrm{MtCO_2e}$)
  - Specific GHG Intensity ($\mathrm{kgCO_2e/t}$ cementitious)
  - Total Water Consumption & Specific Water (m^3/t)
  - Thermal Substitution Rate (TSR %) & WHRS Capacity
- **Data Callout:** Every point has an inline badge: `[STANDALONE]`, `[ASSURED: TUV INDIA]`.

Tab 2: Target Ledger & Trajectory Status

- Table of stated corporate commitments:

Specific Net CO2

Baseline Year / Value	2020: 612 kg/t
Target Year / Value	2030: 465 kg/t
FY26 Actual	521 kg/t
Progress	61.9%
Status	`ON TRAJECTORY`
Verification	SBTi 1.5°C Validated

Renewable Power

Baseline Year / Value	2020: 19.0%
Target Year / Value	2030: 75.0%
FY26 Actual	44.1%
Progress	44.8%
Status	`NEAR TRAJECTORY`
Verification	Assured BRSR Core

Water Positivity

Baseline Year / Value	2020: 3.2x
Target Year / Value	2030: 5.0x
FY26 Actual	4.6x
Progress	77.8%
Status	`ON TRAJECTORY`
Verification	DNV Certified

Tab 3: Claims vs. Evidence Audit

- Exposes corporate ESG claims and verifies empirical grounding:
  - *Claim:* "Targeting 5x Water Positive by 2030."
  - *Alignment Status:* `SUPPORTED BY REGULATORY DATA`.
  - *Evidence:* Groundwater extraction decreased from 1.82 Mm³ (FY23) to 1.45 Mm³ (FY26) while rainwater harvesting in mine pits reached 6.8 Mm³.
  - *Assurance:* Independent water balance verification by TUV India.

Tab 4: Contradiction & Quality Registry

- Shows any detected data discrepancies or historical restatements:
  - *Issue ID:* `Q-JK-2024-01`
  - *Classification:* Historical Restatement.
  - *Description:* FY23 Scope 1 emissions restated from 7.78 Mt in FY23 filing to 7.98 Mt in FY24 filing due to acquisition of Panna grinding unit.
  - *Resolution:* Both versions preserved; restatement note validated.

Tab 5: Provenance Inspector & Source Documents

- Complete register of raw files:
  - `JKCEMENT-2026.xml` (SHA-256: `708938a560c7e0b...`) [View Raw XML] [Download]
  - `JKCEMENT-2025.xml` (SHA-256: `9a5d3adac84a236...`) [View Raw XML] [Download]
  - `JKCEMENT-2024.xml` (SHA-256: `aab9843f51cdb99...`) [View Raw XML] [Download]
  - `JKCEMENT-2023.xml` (SHA-256: `4090fa71c05ec1b...`) [View Raw XML] [Download]

---

3. Click-to-Evidence Modal Dialog

When a user clicks any data point (e.g. Scope 1 value: `12,725,657 tCO2e`), the Traceability Modal renders:

```
┌──────────────────────────────────────────────────────────────────────────┐
│  OBSERVATION PROVENANCE INSPECTOR                                  [X]   │
├──────────────────────────────────────────────────────────────────────────┤
│  Metric: Gross Scope 1 GHG Emissions (AIR-GHG-001)                       │
│  Company: J.K. Cement Limited | Reporting Period: 2025-04-01 to 2026-03-31 │
├──────────────────────────────────────────────────────────────────────────┤
│  NORMALIZED VALUE: 12,725,657.0 tCO2e                                    │
│  ORIGINAL REPORTED VALUE: 12725657                                       │
│  ORIGINAL UNIT MEASURE: Non-SI:tCO2e (Context: DCYMain)                  │
│  EXTRACTION METHOD: Deterministic XBRL Engine (XPath: /*/*[1104])        │
│  DOCUMENT SHA-256: 708938a560c7e0b3667fccb874f4a007f0fe85aacc9d2af51e309... │
│  SOURCE URL: https://nsearchives.nseindia.com/corporate/xbrl/BRSR_532644_... │
├──────────────────────────────────────────────────────────────────────────┤
│  INDEPENDENT ASSURANCE RECORD:                                           │
│  Provider: TUV India Private Limited | Standard: ISAE 3000 / SEBI BRSR Core │
│  Assurance Level: REASONABLE ASSURANCE | Opinion: UNMODIFIED             │
│  Audit Page: Section C Annexure II (Signed: 2026-06-25)                  │
├──────────────────────────────────────────────────────────────────────────┤
│  [ Copy Raw Fact JSON ]  [ View Context XML ]  [ Open Regulatory Source URL ] │
└──────────────────────────────────────────────────────────────────────────┘
```

DOCUMENT 15

AWE Sector Dashboard Specification & UI Wireframe

15-sector-dashboard-wireframe.md

Document ID: AWE-UI-SECT-v1.0
Route: `/sector/cement` (NIC 23941 / 23942 — Cement & Clinker)
Entities Included ($n = 6$): UltraTech Cement, Ambuja Cements, ACC, Shree Cement, Dalmia Bharat, JK Cement.

---

1. Top Sector Distribution Header

```
┌──────────────────────────────────────────────────────────────────────────────────────────┐
│  INDIAN CEMENT SECTOR BENCHMARK OVERVIEW (FY2025-26)                                     │
│  Sample Size: 6 Major Producers | Total Cementitious Capacity Tracked: ~280 MTPA (~65% National Share)  │
├──────────────────────────────────────────────────────────────────────────────────────────┤
│  AWE PERFORMANCE DISTRIBUTION (0 - 100)                                                  │
│  [Min: 64.2] ——— [P25: 69.1] ——— [MEDIAN: 73.8] ——— [P75: 78.4] ——— [Max: 81.2]          │
│                                                                                          │
│  PILLAR MEDIAN SCORES:                                                                   │
│  • AIR: 71.5/100 (45% Wt)  — Median Specific Net CO2: 538 kgCO2/t | Median Renewable Power: 38.5%    │
│  • WATER: 76.2/100 (25% Wt) — Median Specific Water: 0.165 m3/t | Median Water Stressed Withdr: 54%   │
│  • EARTH: 74.0/100 (30% Wt) — Median TSR: 11.2% | Median Clinker Factor: 68.4% | Recycling: 94.5%    │
├──────────────────────────────────────────────────────────────────────────────────────────┤
│  EVIDENCE CONFIDENCE & COVERAGE AUDIT:                                                   │
│  • Sector Mean Evidence Confidence: 91.8% (All 6 filings verified by accredited independent auditors)   │
│  • Sector Mean Score Coverage: 86.4% (Consistently strong reporting on Core Principle 6 indicators)   │
└──────────────────────────────────────────────────────────────────────────────────────────┘
```

2. Sector Peer Comparison Matrix

```
┌──────────────────────────────────────────────────────────────────────────────────────────┐
│  COMPANY COMPARISON TABLE (FY2025-26 STANDALONE/CONSOLIDATED NORMALIZED)                 │
└──────────────────────────────────────────────────────────────────────────────────────────┘
```

Company	Specific CO2 (kgCO2/t)	Renew. Power Share (%)	Specific Wat (m3/tonne)	TSR (AFR) (%)	Clinker Fact (%)	AWE Perf Score	TRS EC Conf.
Dalmia Bharat	481.0	42.5%	0.138	18.2%	61.2%	81.2	94.0%
JK Cement	521.0	44.1%	0.142	14.8%	65.2%	74.2	94.0%
Shree Cement	518.0	53.2%	0.155	12.4%	67.8%	76.5	93.5%
UltraTech	548.0	35.8%	0.172	10.5%	69.5%	71.0	92.0%
Ambuja Cement	532.0	32.0%	0.168	9.8%	68.1%	72.4	91.5%
ACC Limited	554.0	28.5%	0.180	8.2%	70.4%	66.8	86.0%

```
Notes:
* Dalmia Bharat and UltraTech reported on Consolidated basis; JK Cement, Shree, Ambuja, ACC on Standalone.
* ACC Limited is an operating subsidiary of Ambuja Cements (~50.05% ownership); excluded from group sums.
```

3. Boundary Harmonization & Comparability Warnings

The dashboard features an interactive **Data Quality Banner**:

1. **Adani Group Consolidation Warning:** `[CAUTION]` Ambuja Cements and ACC Limited shifted from 15-month consolidated periods in FY22–23 to 12-month standalone periods in FY23–24, FY24–25, and FY25–26. Multi-year growth rates exclude the FY23 baseline to prevent skew.
2. **Denominators Verified:** All specific carbon intensities are benchmarked against *Cementitious Product Outflow* (not Clinker production).
3. **Assurance Scope Confirmed:** All 6 producers in FY26 utilized reasonable assurance under ISAE 3000 / BRSR Core standard from accredited providers (TUV India, SGS India, TUV SUD).

DOCUMENT 16

AWE Automated Research Story Generator & Templates

16-research-story-templates.md

Document ID: AWE-STORY-TEMP-v1.0
Status: Automated environmental visual journalism framework.
Version Anchor: AWE-JOURNALISM-2026-v1.0

---

1. Journalism Engine Philosophy

The AWE Story Engine transforms cold regulatory filings into evidence-grounded visual journalism.

- **Rule of Determinism:** The AI model is strictly prohibited from inventing or approximating statistics. Every number in a published story is populated from a frozen, deterministic database query.
- **Lineage Footnote:** Every story concludes with a machine-readable methodology block containing query hash, data snapshot ID, and document SHA-256 hashes.

---

2. Research Story Archetypes & Templates

Template 1: "The Decarbonisation Divergence" (Growth vs. Efficiency)

Headline:

India’s Cement Giants Decarbonise per Tonne—Yet Absolute Kiln Emissions Rise {pct_abs_growth}%

Key Finding:

Between {start_fy} and {end_fy}, average net carbon intensity among India's six largest cement producers dropped from {baseline_intensity} to {current_intensity} kgCO2e/tonne ({intensity_drop}% improvement). However, due to rapid infrastructure capacity expansion, total annual Scope 1 emissions surged by {abs_surge_mt} million tonnes.

Data Table (Primary Statutory BRSR Filings):

UltraTech Cement

FY22 Scope 1 (Mt)	61.45
FY26 Scope 1 (Mt)	88.88
4-Year Absolute Chg	+44.6%
Specific CO2 (kg/t)	548.0

Ambuja Cements

FY22 Scope 1 (Mt)	15.29*
FY26 Scope 1 (Mt)	22.59
4-Year Absolute Chg	+47.7%
Specific CO2 (kg/t)	532.0

Shree Cement

FY22 Scope 1 (Mt)	18.33*
FY26 Scope 1 (Mt)	22.03
4-Year Absolute Chg	+20.2%
Specific CO2 (kg/t)	518.0

Dalmia Bharat

FY22 Scope 1 (Mt)	12.55
FY26 Scope 1 (Mt)	15.75
4-Year Absolute Chg	+25.5%
Specific CO2 (kg/t)	481.0

JK Cement

FY22 Scope 1 (Mt)	7.78
FY26 Scope 1 (Mt)	12.73
4-Year Absolute Chg	+63.6%
Specific CO2 (kg/t)	521.0

ACC Limited	
FY22 Scope 1 (Mt)	12.80*
FY26 Scope 1 (Mt)	14.10
4-Year Absolute Chg	+10.2%
Specific CO2 (kg/t)	554.0

*Note: Ambuja, Shree, and ACC historical comparatives normalized from FY23/24 filings.

Visual Chart:

[Chart: Viz 01 - Scope 1 Decarbonisation Trajectory with Discontinuity Markers]

The Explainer:

The cement sector represents the physical backbone of India's urbanisation. While process innovations—such as low-clinker blended cements (PPC/PSC), thermal substitution with refuse-derived fuels, and captive waste heat recovery systems—have successfully reduced the emissions required to produce a single bag of cement, aggregate market demand has outpaced unit efficiency gains.

Methodology & Provenance:

- Query Run ID: `q-run-20260924-cement-decarb-01`
- Sources: NSE BRSR Filings (FY22 to FY26) for UltraTech (INE481G01011), Ambuja (INE079A01024), ACC (INE012A01025), Shree (INE070A01015), Dalmia (INE00R701025), JK Cement (INE823G01014).
- Assurance Status: FY26 filings covered by Independent Reasonable Assurance under SEBI BRSR Core.

Template 2: "Water Reality in Parched Basins" (Water Stress Overlap)

Headline:

Over {stressed_pct}% of Industrial Groundwater Withdrawal by Top Cement Plants Occurs in Critical Water-Stressed Basins

Key Finding:

In FY2025-26, the six tracked cement producers extracted {total_groundwater_mm3} million cubic meters of groundwater. Over half of this extraction was concentrated in over-exploited and critical hydrological assessment units in Rajasthan, Gujarat, and Madhya Pradesh.

Visual Chart:

[Chart: Viz 05 - Specific Water Consumption & Stress Overlap Matrix]

The Explainer:

While companies actively disclose corporate-wide "water positivity" ratios based on artificial recharge structures and mine-pit reservoirs, local aquifer replenishment varies dramatically across operating sites. AWE's facility-level mapping reveals that site-level groundwater stress remains acute despite high corporate-level positivity claims.

Template 3: "Circular Economy: The Kiln Waste Sink" (Industrial Co-Processing)

Headline:

Beyond Landfills: Indian Cement Kilns Co-Process {total_afr_mt} Million Tonnes of Hazardous & Municipal Waste

Key Finding:

Cement kilns operate as high-temperature thermal destructors. In FY26, the sector reached an average Thermal Substitution Rate (TSR) of {mean_tsr}%, consuming {plastic_waste_tonnes} tonnes of municipal plastic waste and industrial fly-ash/slag.

---

3. Social Media & Syndication Formats

1. **LinkedIn Carousel Format:**
   - Slide 1: High-contrast title card with headline and primary metric callout.
   - Slide 2: The Data Gap (e.g. Absolute emissions vs. Specific intensity).
   - Slide 3: Peer Comparison Chart with company rankings.
   - Slide 4: Assurance & Data Quality Badge (highlighting independent audit).
   - Slide 5: Direct citation link to AWE Evidence Explorer.
2. **Research Briefing (PDF Export):**
   - 2-page executive brief formatted for institutional investors, credit rating agencies, and environmental regulators.

DOCUMENT 17

AWE Methodology Versioning & Reproducibility Framework

17-methodology-versioning.md

Document ID: AWE-METH-VERS-v1.0
Status: Canonical governance & data architecture standard.
Version Anchor: AWE-VERSIONING-2026-v1.0

---

1. The Immutability Mandate

In legacy ESG rating schemes, methodology alterations are retroactively applied to past years, silently rewriting historical company performance and obscuring historical ratings drift.

AWE establishes an inviolable reproducibility principle:

> [!IMPORTANT] **Cardinal Rule of Versioning:** Once an AWE Performance Score or Benchmark is published under a specific methodology version (e.g. `AWE-CEMENT-2026-v1.0`), that score is **cryptographically frozen**. It will never be modified, deleted, or overwritten in place.

---

2. Methodology Version String Specification

Methodology identifiers follow a strict semantic syntax: `AWE-{SECTOR}-{RELEASE_YEAR}-v{MAJOR}.{MINOR}`

Examples:

- `AWE-CEMENT-2026-v1.0`: Initial cement methodology release based on SEBI BRSR Core.
- `AWE-CEMENT-2027-v2.0`: Major revision incorporating mandatory value chain Scope 3 disclosures.
- `AWE-STEEL-2026-v1.0`: Specific methodology for primary iron and steel manufacturing.

Components Locked by Methodology Version:

1. **Indicator Materiality Weights:** Exact weights for Air (w_A), Water (w_W), and Earth (w_E).
2. **Sub-Indicator Decomposition:** Formulas for Intensity, Trajectory, Target Progress, and Operational Outcomes.
3. **Canonical Normalization Rules:** Unit conversions, denominator definitions, and physical boundary scopes.
4. **Benchmark Cohort Algorithm:** Mathematical definition of percentiles, Winsorization limits, and outlier thresholds.
5. **TRS Evidence Formula:** Evidence Confidence dimension weights and assurance penalty multipliers.

---

3. Side-by-Side Recalculation (Dual-Horizon Reporting)

When methodology evolves (e.g. from `v1.0` to `v2.0`), AWE supports **Dual-Horizon Recalculation**:

- A company's historical FY2025 performance is retained as:
  - `AWE Performance (FY25 under v1.0): 74.2`
  - `AWE Performance (FY25 recalculated under v2.0): 71.8`
- Both records remain queryable in the API and UI. A visual toggle allows users to compare:
  1. *Historical Reality:* The score calculated at the time of publication under the contemporary methodology.
  2. *Normalized Comparable Time Series:* The entire 5-year history computed under the latest uniform methodology version (`v2.0`).

---

4. Cryptographic Calculation Run Record

Every score calculation produces an immutable `calculation_run` record:

```
{
  "run_id": "8f3e2a1b-9c8d-4e7f-6a5b-0c1d2e3f4a5b",
  "methodology_version": "AWE-CEMENT-2026-v1.0",
  "executed_at": "2026-09-24T18:00:00Z",
  "git_commit_sha": "a1b2c3d4e5f67890123456789abcdef012345678",
  "dependency_lock_hash": "e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855",
  "input_dataset_snapshot_id": "snap-cement-26-xml-filings-20260924",
  "random_seed": 42,
  "parameters_hash": "4a5b6c7d8e9f0a1b2c3d4e5f6a7b8c9d0e1f2a3b4c5d6e7f8a9b0c1d2e3f4a5b"
}
```

Any third-party researcher with access to the raw input snapshot and source code repository can re-run the pipeline and obtain **bit-exact identical scores**.

DOCUMENT 18

AWE Public Governance & Scientific Integrity Framework

18-governance-framework.md

Document ID: AWE-GOV-FRAME-v1.0
Status: Canonical governance charter.
Version Anchor: AWE-GOVERNANCE-2026-v1.0

1. Core Principles of Scientific Independence

To protect against conflicts of interest that have compromised commercial ESG rating agencies:

1. **No Pay-for-Play:** No rated entity, sponsor, or industry association may purchase a higher score, altered benchmark, or privileged rating tier.
2. **Strict Consulting Firewall:** AWE and its operating entity shall never offer paid advisory, assurance, or consulting services to rated entities.
3. **Public Algorithmic Transparency:** The entire mathematical scoring specification, indicator dictionaries, normalization rules, and benchmark formulas must be publicly disclosed under open academic licenses. Any researcher must be capable of independently reproducing AWE Scores.
4. **No Opaque Proprietary "Black Boxes":** Under no circumstances shall proprietary neural network weights or opaque LLM sentiments determine a company's final environmental performance score.

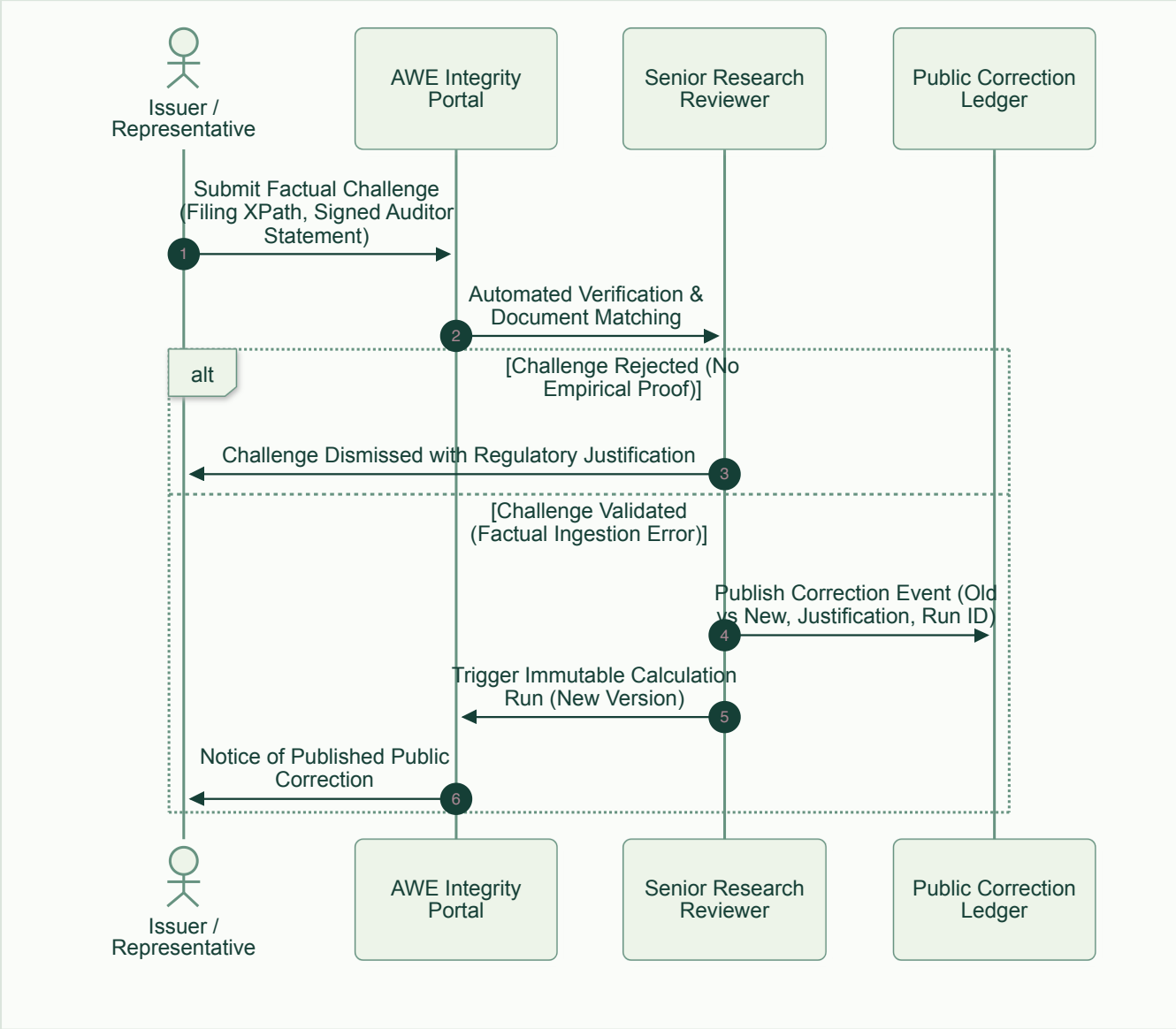
2. Taxonomical Data Lineage Classifications

Every public data point published by AWE must carry one of five strict data lineage labels:

Classification	Definition	Example
REPORTED	Fact extracted directly from an official corporate regulatory filing without arithmetic modification.	Gross Scope 1 GHG emissions of 12,725,657 tCO2e as disclosed in JK Cement FY26 BRSR.
VERIFIED	Fact corroborated by an independent accredited assurance provider under an auditable standard.	Scope 1 & 2 emissions verified under ISAE 3000 Reasonable Assurance by TUV India.
DERIVED	Deterministic mathematical calculation computed from reported facts using public formulas.	Carbon intensity per tonne of cementitious product (Gross Scope 1 ÷ Cement Output).
ESTIMATED	Conservative engineering calculation used when raw values are reported in compound or non-standard units.	Thermal heat consumption estimated from disclosed fuel consumption and net calorific values.
MODELED	Statistical projection or trajectory extrapolation clearly designated as hypothetical.	Projected 2030 emissions trajectory based on historical 3-year CAGR.

3. Transparent Factual Correction Lifecycle

AWE provides rated companies with a structured, transparent mechanism to challenge empirical data errors:



Protocol Standards:

- Challenges must reference statutory public documents. Private corporate letters or confidential internal spreadsheets cannot alter public scores.
- Every successful correction is permanently recorded in the **Public Correction Ledger** stating:
  - Entity Name & Identifier
  - Metric Affected
  - Prior Value & Corrected Value
  - Nature of Error (Parser error, unit scaling error, issuer restatement)
  - Reviewer ID and Timestamp.

4. Independent Methodology Review Board

The AWE Methodology Review Board is comprised of independent academic researchers, environmental scientists, and remote sensing specialists.

- **Review Cycle:** Annual public consultation for all sector indicator dictionaries.
- **Quorum & Independence:** Board members must sign annual conflict-of-interest disclosures confirming zero financial ties to rated industrial corporations.

DOCUMENT 19

AWE cement MVP — implementation recommendation

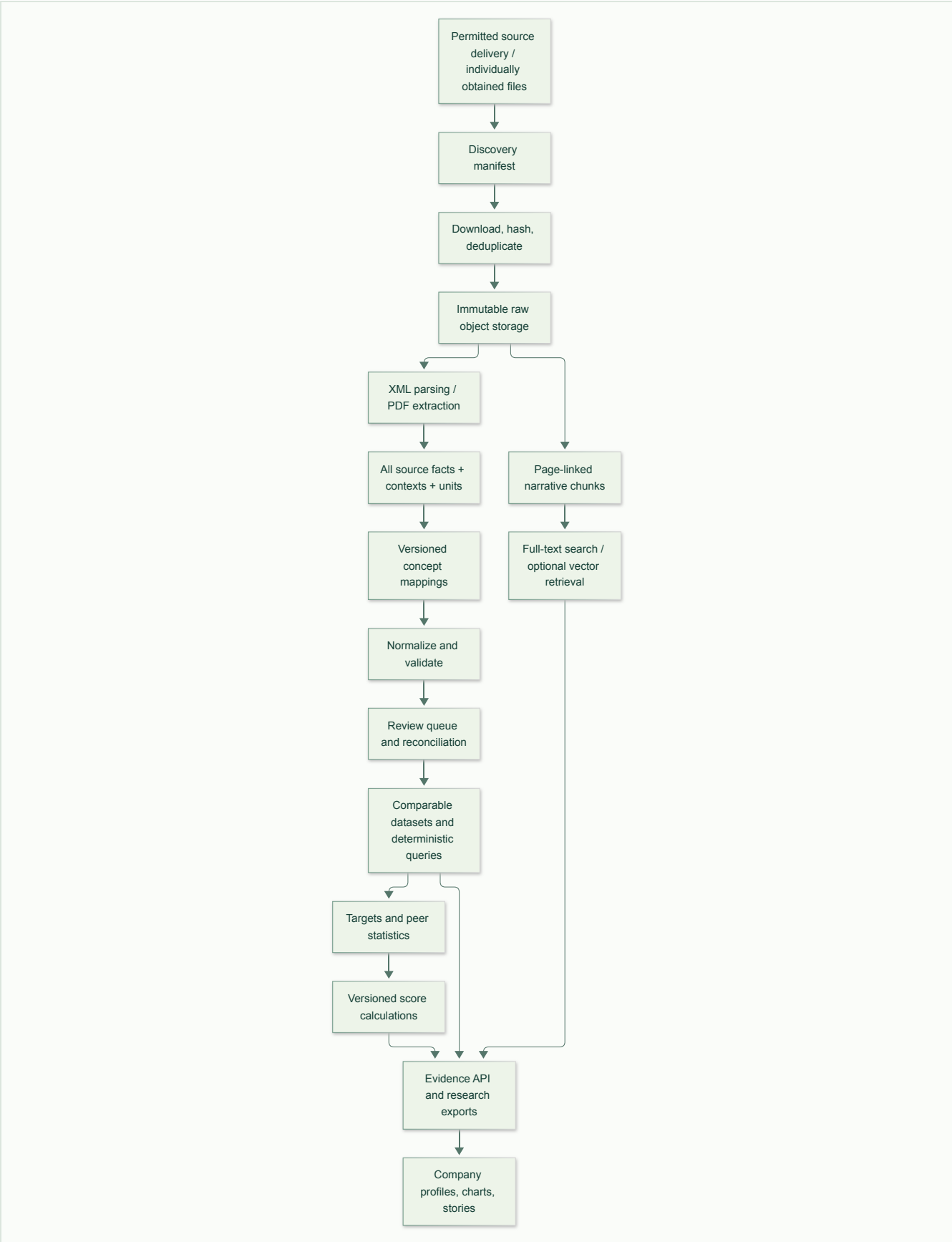
19-mvp-technical-roadmap.md

Status: INFERRED proposal, sequenced after source investigation and taxonomy inspection. No website work has begun. Dates and budgets are not committed.

Recommended initial stack

PostgreSQL for normalized relations, JSONB for preserved XBRL dimensions, and native full-text search for narrative discovery. Use Python workers for XML/PDF parsing and Decimal calculations, S3-compatible object storage with versioning/retention controls, FastAPI for evidence APIs, and a database-backed job table initially. Add pgvector only when retrieval evaluation demonstrates value; add TimescaleDB only if measured workloads justify it. Six annual-report histories do not require a streaming platform or separate graph database.

Next.js and an accessible chart library can follow the data acceptance gate. Start with a read-only internal evidence explorer, then company/sector views. No public composite scores or league tables until methodology, scope and coverage are reviewed.



Milestones and acceptance gates

1. **Finish source-access and census verification.** Account for each of six companies × FY2021–22 through FY2025–26 as downloaded, listed-only, not found, or retrieval-failed. A missing search result is

not proof of absence. Verify BSE independently where NSE gaps exist. Establish permitted bulk/reuse route before scaling. Preserve filing IDs, amendments, raw bytes and hash manifests.

2. **Approve taxonomy mappings.** Retain the complete five-version inventory. Review the provisional Air/Water/Earth classification, custom units, all environmental Core mappings, dimensional relationships and semantic change crosswalk. Every mapped fact resolves to an exact concept version; unmapped facts are preserved.
3. **Build ingestion vertical slice.** Import raw files, identify company and period from content, parse all XBRL facts without an LLM, and persist provenance. Parse one PDF-only historical report as a separate reviewed path. A rerun creates zero duplicate facts. Retries are bounded with exponential backoff; failures enter a queue with source, step, attempt and error. Honour publisher rate/permission requirements.
4. **Normalize and validate.** Test kl→m³, tonne scales, GJ/MWh, compound intensities and percentage representation using source-grounded fixtures. Preserve source value/unit. Flag nil/blank/zero distinctions, unknown units, arithmetic mismatches, boundary changes and unexplained revisions. No silent fixes.
5. **Cement comparability.** Separate clinker, cement, cementitious products and sales-volume denominators; gross/net Scope 1; market/location-based Scope 2; thermal substitution, renewable electricity and renewable total energy. Record captive power, acquired facilities, white/grey cement and geographic coverage. Avoid group overlap. A six-company peer sample is descriptive and sensitive to outliers; do not imply national representativeness.
6. **Target and evidence ledgers.** Reproduce progress formulas from the brief with consistent scope, baseline and denominator. Zero baseline-to-target distance or missing observations returns insufficient data. Store unbounded progress; no arbitrary clamping. Annual trajectory method, tolerance and time basis must be versioned. Independently review assurance engagement scope and claim-evidence alignment.
7. **Benchmarks and candidate methodology.** Publish medians/quartiles only for compatible cohorts, with method (e.g. explicit quantile algorithm), n, exclusions and sensitivity. Candidate sector weights remain draft. Require approved minimum coverage by pillar before publishing a composite. Freeze benchmark membership and method versions so historical results reproduce.
8. **Evidence API and internal explorer.** Read-only company, report, observation, target, benchmark and provenance endpoints. Pagination and filtering are explicit. Exports contain source IDs, units, boundaries and methodology versions. Test authorization for restricted raw files and parameterized queries.
9. **Ten historical visualizations.** Build only those supported by data: Scope 1, Scope 2, comparable combined intensity, energy mix, water consumption, water intensity, waste generation, waste recovery, target progress and coverage/assurance timeline. Each needs definition, unit, period, n, source, method and comparability caveats. Break lines at material boundary changes. Narrative numbers are taken from frozen deterministic query results.
10. **Public release review.** Independent methodology review; correction mechanism; commercial-conflict policy; evidence reproduction audit; data rights review; accessibility review. Release evidence profiles before rankings if scoring is not ready.

Research interface

An LLM may translate questions into a constrained query representation. Validate against an allowlisted semantic layer, compile parameterized read-only SQL, set row/time limits, and return query/run IDs plus evidence-backed results. The model explains returned data and states missingness. Never allow arbitrary database writes or quantitative answers from memory. RAG chunks retain company, report version, exact period, page/section, source URL and hash. Evaluate retrieval on assurance exclusions, boundary changes, target baselines and contradictory facts before release.

Scale gate

Only expand to 1,000+ companies after repeat ingestion, version migration, evidence resolution, rights coverage and review throughput work for the cement slice. Partitioning, additional queues and vector indexes follow measured scale. International expansion adds jurisdiction-specific taxonomy mappings while preserving the same assertion/provenance model.

Remaining deliverables

The brief's detailed TRS methodology, sector score calibration, cement indicator dictionary, executable database migrations, quality rule suite, API specification, visualization library, wireframes, story engine and working prototype are later milestones. The current package does not pretend these are complete. The first task's remaining verification gaps are enumerated in the research README and coverage register.