

Yemen Economic Transparency Observatory

A constitutional-level technical blueprint for evidence-based economic intelligence in Yemen. This system transforms fragmented data into decision-grade analysis through rigorous provenance tracking, split-system architecture, and zero-tolerance for fabrication.





CONTEXT

Yemen's Fragmented Political Economy

Yemen's economy operates under divided authorities—Aden's internationally recognised government and Sana'a's de facto authorities. Data sources conflict. Humanitarian and economic systems overlap. Trust deficits pervade every metric. Traditional data platforms cannot navigate this complexity without conflating incompatible figures or hiding contradictions.

YETO addresses this reality head-on: it treats division as a feature, not a bug, tracking each regime separately unless users explicitly choose comparison.



The Intelligence Gap



Fragmented Sources

Conflicting Data

Trust Deficit

Poor Decisions

Decision-makers face an impossible choice: rely on incomplete international datasets, trust partisan local sources, or operate blind. YETO closes this gap by creating a single evidence store that preserves contradictions, tracks provenance, and refuses to invent answers.



MISSION

What YETO Does

Ingests Continuously

Pulls data from international databases, UN agencies, Yemeni authorities, humanitarian systems, and research bodies—2010 to present, updating daily.

Enforces Provenance

Every number carries source metadata: publisher, dataset name, retrieval date, license terms, confidence rating, transformations applied.

Produces Intelligence

Delivers dashboards, AI-powered analysis, automated reports, and scenario simulations—all grounded strictly in stored evidence.

Maintains Transparency

Users can trace any claim back to source documents. Contradictions are displayed, not hidden. Uncertainty is labelled explicitly.

Absolute Rules Governing YETO

These twelve principles are non-negotiable constraints enforced in code, UI, and AI logic. They form the constitutional framework of the platform.

O1

No Hallucination

Never invent data. If unavailable, display "Not available yet" and trigger a gap ticket.

O2

Every Number Has a Home

Each statistic shows source, URL, dates, license, geographic scope, units, confidence, transformations.

O3

Triangulation for High-Stakes Outputs

For sensitive indicators, seek multiple sources. Display conflicts side by side.

O4

Versioning

Preserve all data updates. Enable "what was known when" reconstruction.

O5

Do No Harm

No PII exposure. No guidance enabling violence, sanctions evasion, or market manipulation.

O6

Arabic-First and Accessible

Full RTL support, bilingual glossary, WCAG 2.1 AA compliance.

O7

Licensing Compliance

Respect robots.txt, honour paywalls, store only metadata for restricted content.

O8

Tested CI/CD

All changes pass automated tests. Main branch protected; changes require pull requests.

O9

Single Source of Truth

All displays use YETO's internal evidence store only. Full reproducibility.

O10

Public Corrections Policy

Visible "Report an Issue" on every chart. Corrections published transparently.

O11

Split-System Enforcement

Tag data by regime. Never conflate Aden vs Sana'a metrics without user action.

O12

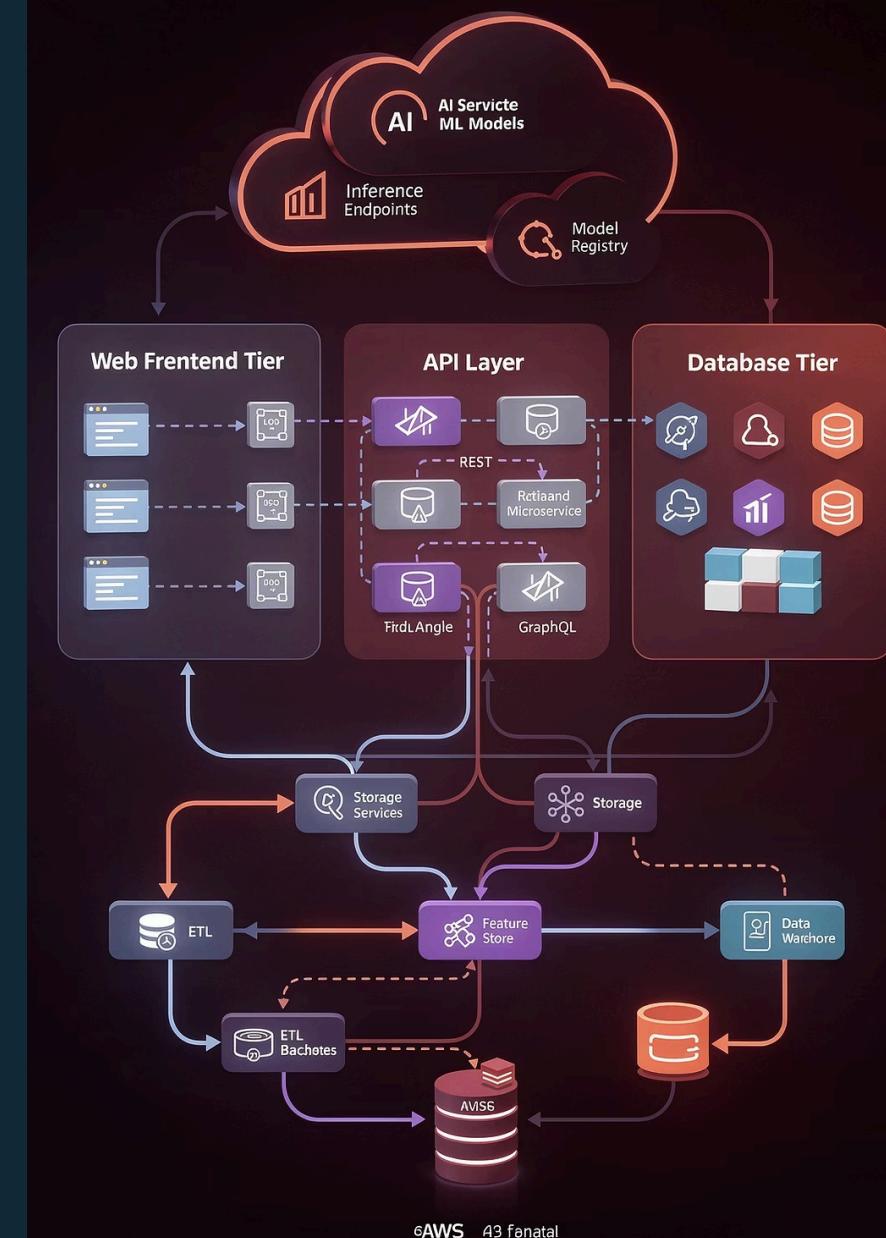
Evidence Packs

High-stakes outputs include supporting sources, contradictions, confidence ratings.



System Architecture at a Glance

YETO is a cloud-native platform deployed on AWS. A React/Next.js frontend delivers bilingual UI. FastAPI or Node.js backend orchestrates data ingestion, search, and AI. PostgreSQL stores structured data, S3 holds documents, OpenSearch enables full-text retrieval. The AI assistant uses retrieval-augmented generation to ground answers in evidence.



Data Universe: What Enters the Platform

YETO integrates hundreds of sources spanning international institutions, UN agencies, Yemeni authorities (both sides), humanitarian datasets, research bodies, and media for event detection. Coverage runs from 2010 to present, continuously updated.



Multilateral Data APIs

World Bank, IMF, UN Comtrade, FAO, ILO, WHO, UNICEF—accessed via RESTful endpoints.



Humanitarian Data

OCHA FTS, HDX HAPI, UNHCR, IOM DTM, ReliefWeb—funding flows, displacement, food security.



Conflict Data

ACLED, UCDP—battle deaths, incident locations, economic disruption events (licensed access).



Yemeni Institutions

Central Bank (Aden & Sana'a), Ministry of Finance, CSO, port authorities—official bulletins, exchange rates.



Development Programmes

SFD, PWP, Yemen Microfinance Network—project outputs, community-level data.



Research & Think Tanks

Sana'a Center, Rethinking Yemen's Economy, Chatham House, Crisis Group—PDF reports, policy briefs.

Source Management Architecture

Source Registry

Master table cataloguing every data source: name, type, credibility rating, update frequency, access method (API, scraper, manual), license terms, sector tags. Pre-seeded with major known sources at initialization.

Discovery Engine

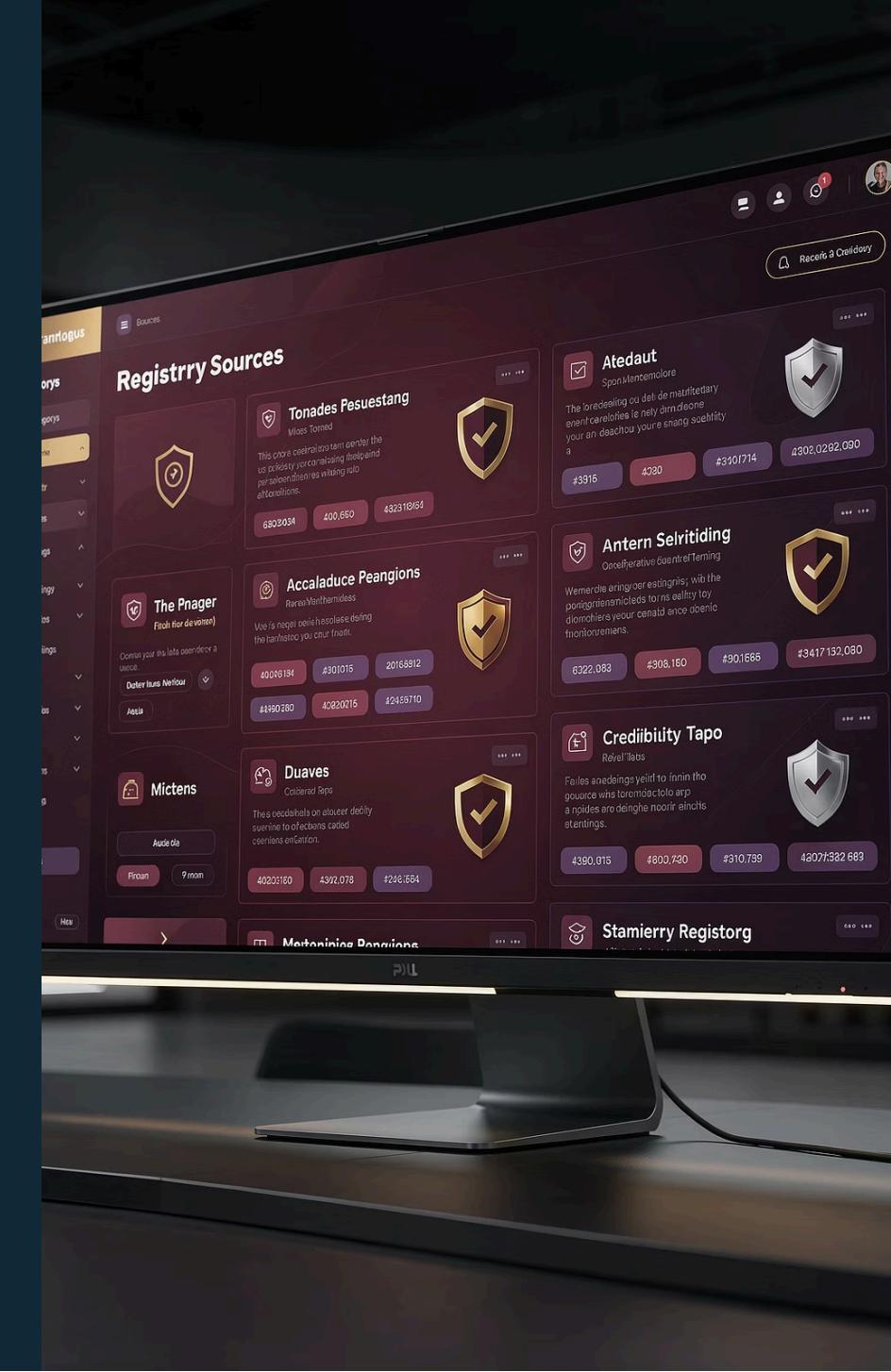
Weekly jobs query OpenAlex for scholarly articles, ReliefWeb for reports, HDX for datasets, IATI for projects. Discovered candidates enter a Review Queue with automated metadata suggestions.

Partnership Portals

Trusted partners upload CSV files to secure S3 buckets. Contributors receive credentials for direct submission via portal interface.

Licensing Enforcement

Registry notes open vs restricted sources. For restricted content, YETO logs gaps and contacts providers. No unauthorized ingestion—compliance with R6 is absolute.



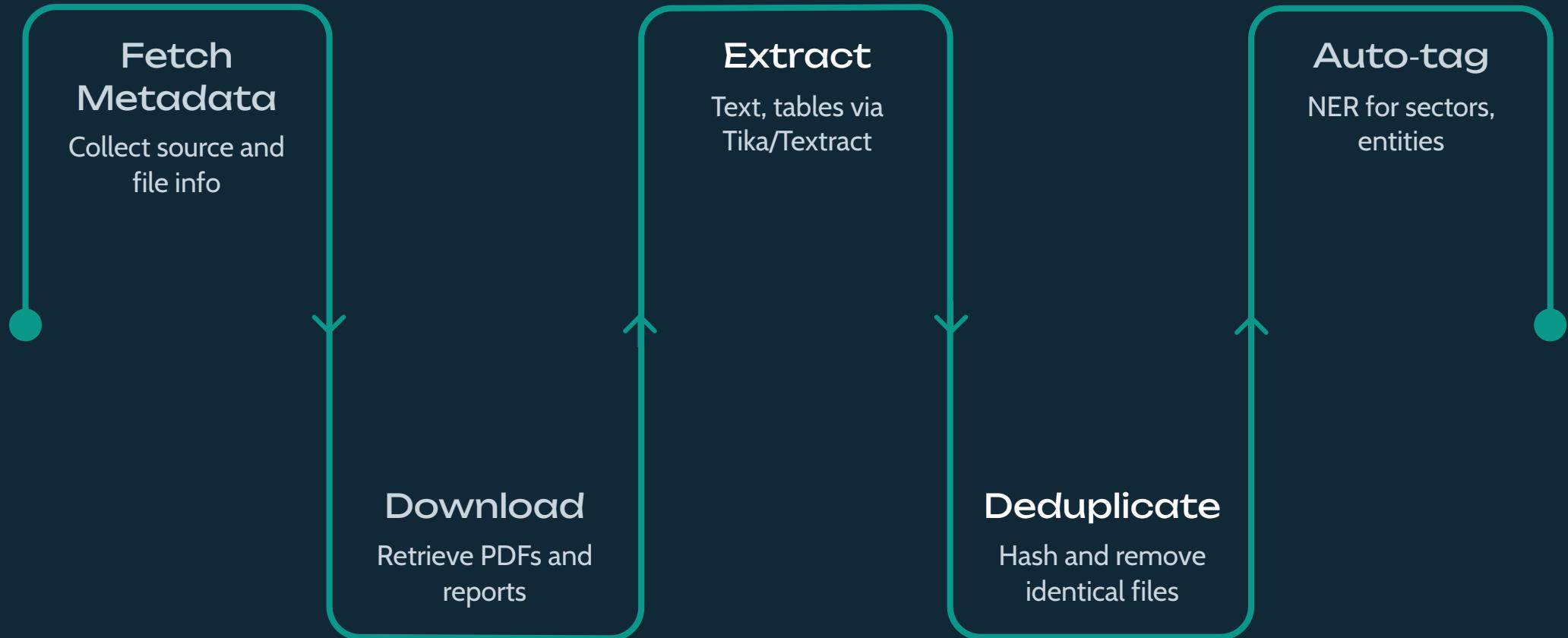


Data Model: The Knowledge Graph Core

PostgreSQL schema connects sources, documents, series, observations, geographies, entities, projects, events, and provenance. Every entity links to others, forming a traversable graph that powers AI retrieval and related content suggestions.

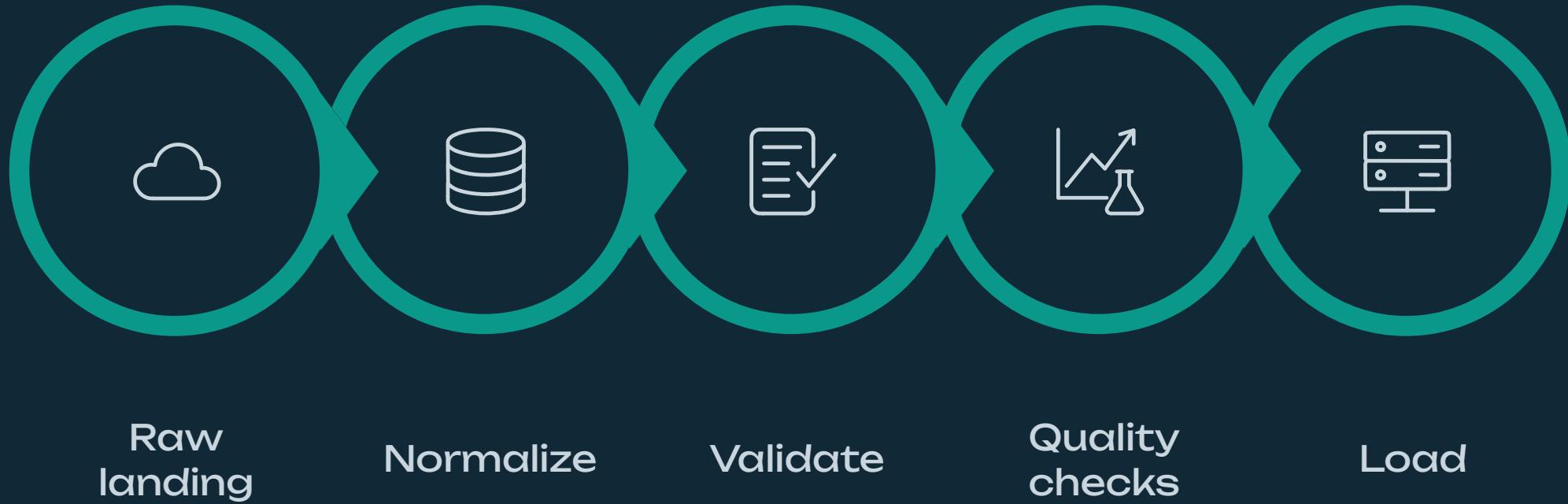
- Sources**
Master registry of data providers with credibility ratings, update cadence, licensing.
- Documents**
PDFs, reports, HTML pages with text extraction, translation status, thematic tags.
- Series**
Time series definitions: name, units, frequency, source, geographic level, regime tag.
- Observations**
Fact table for numeric data points with series link, date, value, vintage, confidence.
- Geographies**
Country, governorates, districts with GIS coordinates and regime control flags.
- Entities**
Institutions, ministries, banks, NGOs, donors—relationships tagged across content.
- Projects**
Aid/development initiatives with budgets, locations, implementers, status tracking.
- Events**
Policy changes, conflict incidents, economic shocks—linked to data series impacts.

Ingestion Pipeline: Document Workflow



PDFs and reports enter via APIs or uploads. Apache Tika or AWS Textract extracts text. Deduplication via hashing prevents storage of identical files. NER tags sectors, geographies, entities. Embedding models create vectors for semantic search. Translation engine (AWS Translate + glossary enforcement) produces bilingual content. OpenSearch indexes both languages.

Ingestion Pipeline: Structured Data Workflow



APIs return JSON or CSV. Raw files land in S3 with organized paths. Parsing scripts normalize to canonical schema. Validation checks required fields and types. QA detects outliers (statistical deviation), consistency issues (gaps, duplicates), and contradictions (conflicts between sources). Passing data loads into Observations table with provenance logging. Post-load triggers recompute derived indicators, update search indexes, evaluate alerts, refresh coverage scorecard.

Quality Assurance Mechanisms



Outlier Detection

Statistical methods flag values deviating strongly from historical trends or plausible ranges.



Consistency Checks

Verify time continuity, ensure disaggregated data sums match totals, validate regime tags.



Contradiction Detection

Compare values across sources for same indicator/date. Store conflicts in Contradictions table.

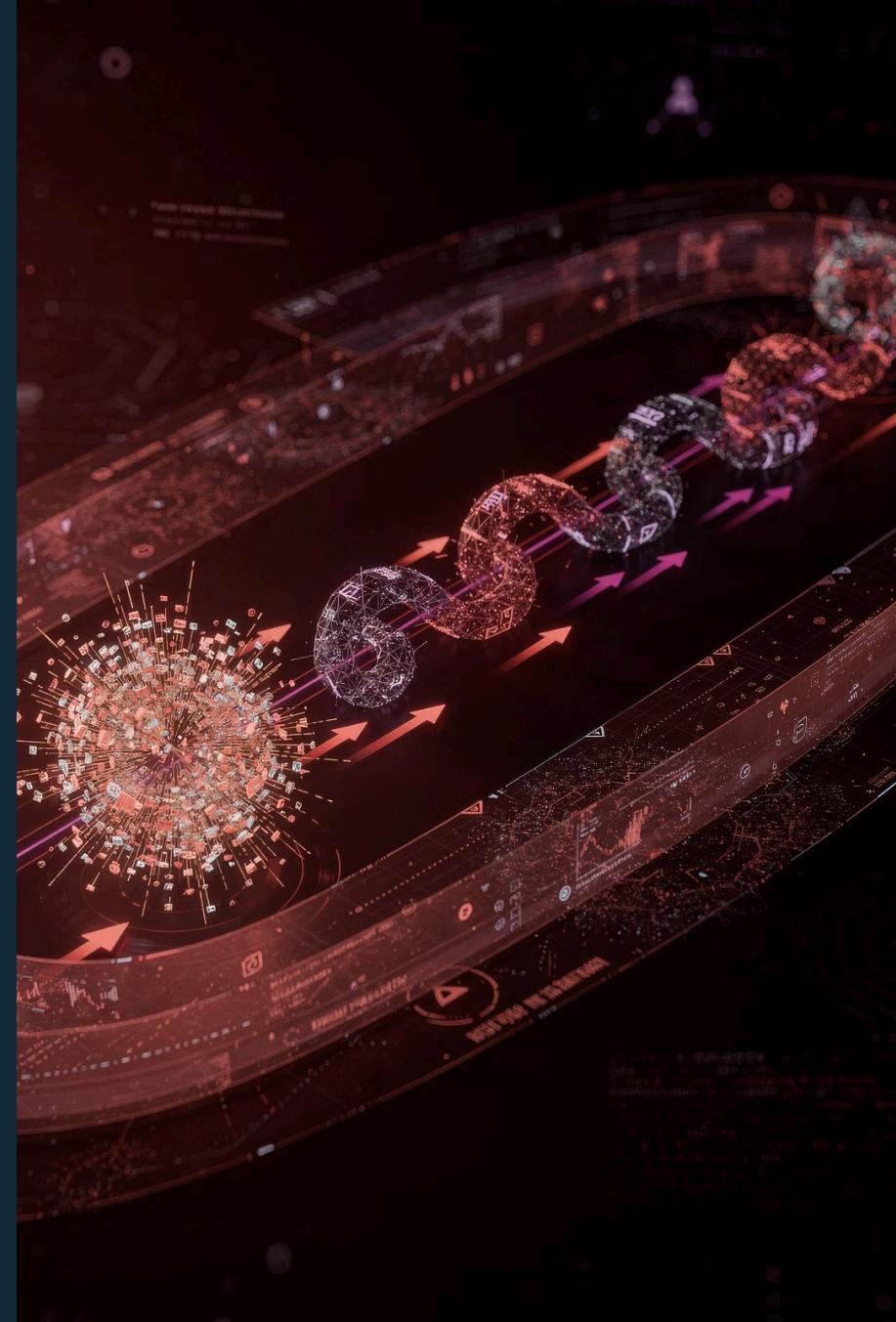


Human Review Queue

Admin portal highlights flagged anomalies for analyst verification and resolution.

Provenance and Trust Architecture

Every data point carries complete lineage: publisher name, dataset identifier, publication date, retrieval timestamp, license terms, geographic scope, time coverage, definitions, confidence rating, transformations applied. This "source pack" appears in UI metadata panels. Exports embed citations automatically.



Confidence Rating System

Grade A: Verified

Official statistics from recognised authorities, thoroughly cross-checked, consistent across sources.

Grade C: Proxy/Modeled

Estimates or proxies used in absence of direct data. Collected under challenging circumstances.

Grade B: Credible

Generally reliable but possibly incomplete, lagged, or from single well-regarded source.

Grade D: Disputed

Known conflicts between sources or low reliability due to methodology. Shown with warnings.

Confidence grades appear in UI with icons. Grade C/D data includes explanatory tooltips and may trigger evidence pack requirements for any analysis using them.

Contradiction Handling: Disagreement Mode

When YETO finds conflicting values for the same indicator and date, it stores both in the Contradictions table. Default view shows the higher-confidence source but indicates another exists. Users can enable "disagreement mode" to overlay both lines on charts with annotations explaining differences—different base years, inclusion of estimates, contested sectors.

The platform treats users as capable of understanding nuance. It presents multiple perspectives rather than hiding discrepancies or averaging incompatible figures.

Example: GDP 2014

- IMF: \$35.6B (includes estimates for northern regions)
- World Bank: \$33.2B (excludes contested areas)
- YETO shows both with explanations, lets user choose or compare





Versioning: What Was Known When

YETO never silently overwrites data. When sources revise figures, new observations are added with vintage identifiers. Old values remain accessible via version history. Users can reconstruct "as of 2022, what did World Bank report for 2020 GDP?" vs "current revised estimate."

Changelog pages compile significant revisions with timestamps and reasons. This enables audit trails and research reproducibility.



Split-System Data Enforcement

Every dataset, event, or document tagged with regime indicator: Aden/IRG, Sana'a/de facto, or mixed/unknown. UI never conflates data from divided authorities. Default view keeps them separate. Comparisons only shown if user explicitly enables compare mode with warnings.

Example: Central Bank currency data appears as "CBY Aden exchange rate" vs "CBY Sana'a exchange rate"—distinct series, never merged into single "Yemen rate" without user action.

Derived Analytics and Calculations



Growth Rates

Year-over-year and month-over-month changes computed from base series with formulas documented.



Financial Indicators

Parallel vs official exchange rate spreads, regional price divergences—formulae exposed.

All derived metrics stored as separate series flagged as computed. Model registry links them to inputs and formulas. Users see how calculations were performed.



Index Conversions

Rebasing CPI, converting nominal to real values using deflators—all transformations logged.



Composite Indices

Weighted combinations (stress indices) with transparent methodology and adjustable weights.



Anomaly Detection and Alerting

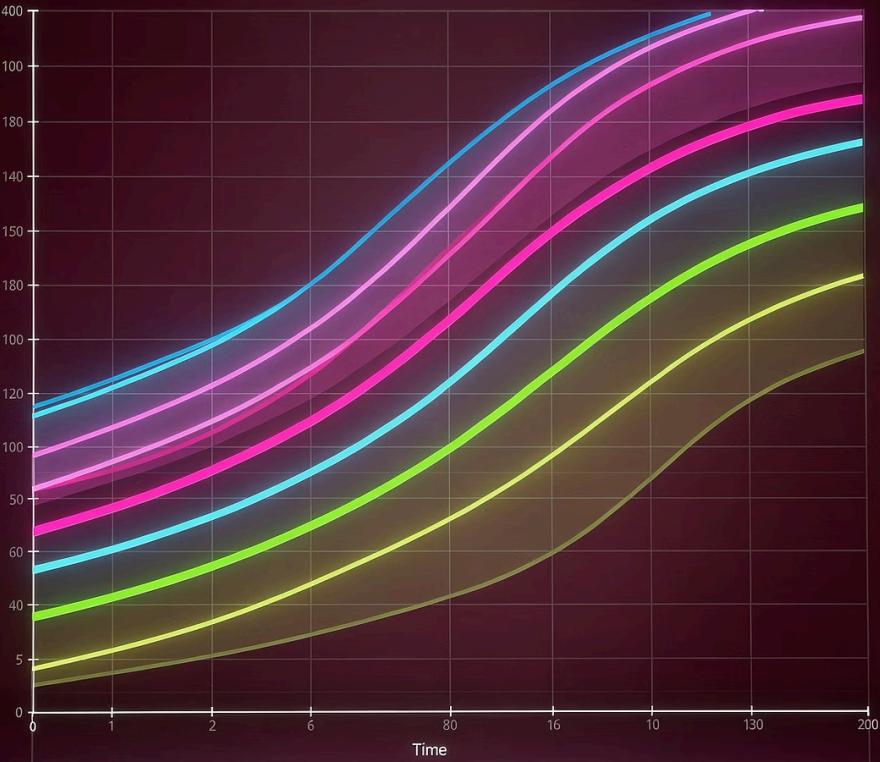
Rule-Based Triggers

Simple thresholds set by system or users: "Inflation exceeds 50%", "Fuel price increases more than 20% in a week". Transparent and domain-informed.

Triggered alerts generate Evidence Packs: what rule fired, which data points, confidence of signal, hypotheses on drivers, historical analogs, suggestions for further monitoring. Users subscribe to custom alerts by topic or threshold.

Statistical Methods

Rolling z-scores, Isolation Forest, ARIMA residuals detect unexpected deviations without predefined thresholds. Signals include significance measures.



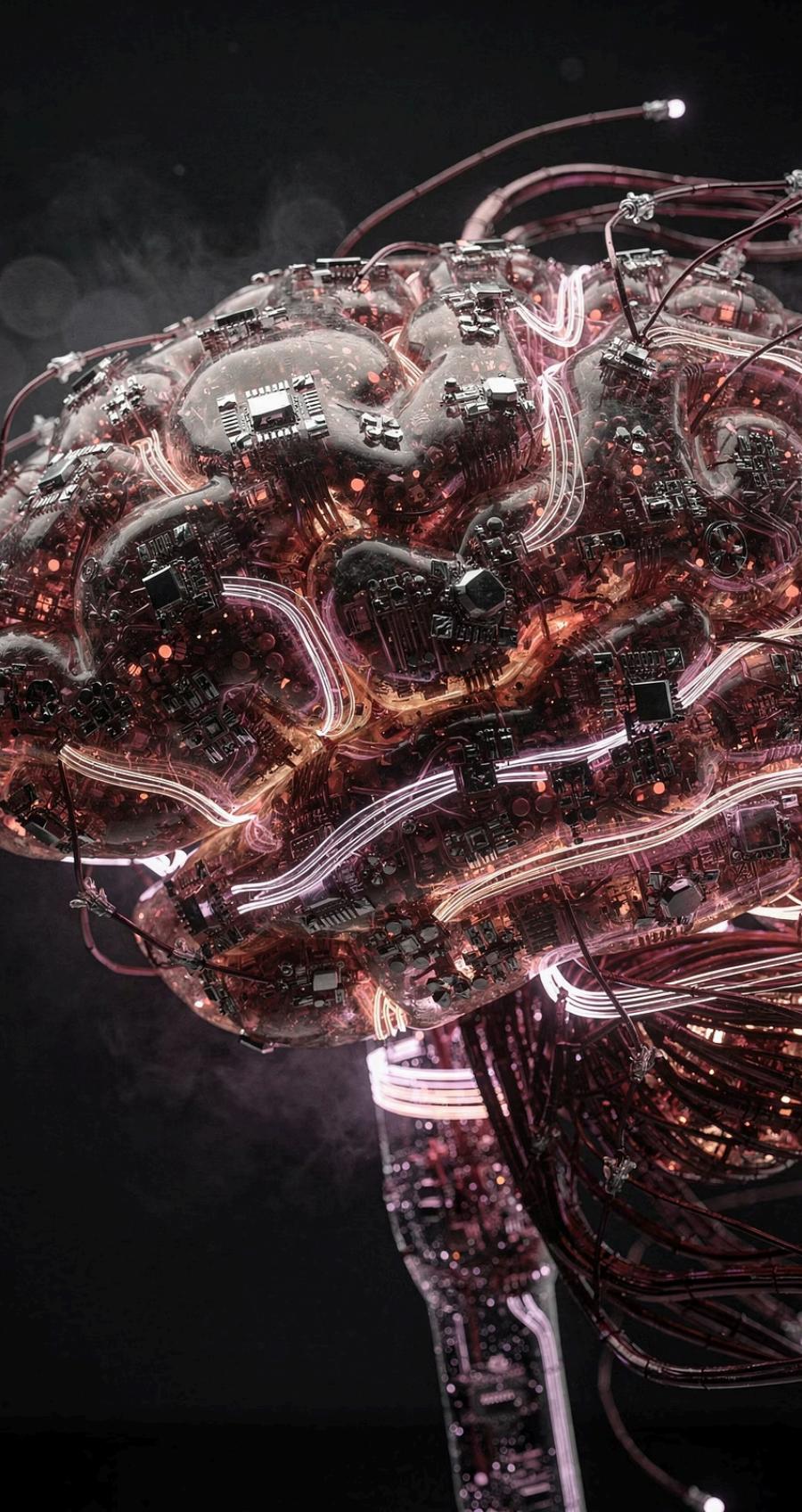
Forecasting and Scenario Simulation

YETO offers cautious forecasting via transparent time-series models (ARIMA, exponential smoothing) with uncertainty ranges. More complex ML models introduced only after backtesting proves value, accompanied by Model Cards explaining assumptions and limitations.

Scenario Simulator allows "what if" explorations: adjust oil prices, port throughput, aid flows. Semi-structural economic model with explicit elasticities. Assumptions not evidence-backed are labelled clearly. Outputs include sensitivity analysis showing how results vary with different parameters.

"One Brain" Intelligence Engine

The YETO Economic Intelligence Assistant is an AI-powered conversational interface using Retrieval-Augmented Generation. It answers questions by retrieving relevant data and documents from YETO's evidence store, then generating structured responses grounded strictly in that context—no hallucination permitted.



AI Assistant: Structured Answer Format

Answer

orem ipsum dolor sitt aisten
mesagees coct ognt eedts timent
heer et. Tresestioms thert coret
opragest he itunes ecauges them
necostost, Tedeved, heft ecopent
opeme deruces dduefeut.

Evidence

orem ipsum dolor sitt aisten
mesageess coct ognt eadts ttinent
heer et. Tresestioms thert coret
opragest he Itunes ecauges them
necostost, Tedeved, heft ecopent
opeme deruces dduefeut.

Key Takeaways

orem ipsum dolor sitt aisten
theragest, cotcoront eadis timent
thearet. Ineusstumst thot doret
conpages ihe ffumes eceruges then
thecedtoot, teeovet., hee coopent
coperre deruces didusfeut.

Confidence

orem ipsum dolor siit aisten
theragest, cotcoront eadis timent
thearet. Ineusstumst thot doret
conpages the fomes ecaruges then
thoeedtool, teeovet, hee coopent
toperre dernces diduefeut.

O1

Answer

One-paragraph direct response

O2

Key Takeaways

3–7 bullet points with citations

O3

Evidence

Data values, charts, quotes with sources

O4

What Changed

Timeline of evolution and events

O5

Why It Matters

Implications and significance

O6

Drivers

Ranked factors with evidence links

O7

Options

2–5 policy actions with tradeoffs

O8

Uncertainty

Candid limitations and blind spots

O9

Next Questions

Suggestions for deeper inquiry

Each answer includes "Show Evidence Pack" button revealing underlying sources, datasets, report excerpts—full transparency on how conclusions were drawn.



Agent Modes: Contextual Personas

Citizen Explainer

Simple language, accessible
Arabic/English, no jargon

Policymaker Brief

Formal, policy-focused,
recommendation-oriented

Donor Analyst

Aid data emphasis,
transparency, accountability
outcomes

Bank Compliance

Financial regulations,
sanctions, compliance issues

Research Librarian

Thorough citations, academic rigor, detailed
references

Data Steward

Provenance details, quality checks, gap
identification

Scenario Modeler

What-if guidance, assumption setting,
sensitivity analysis

Mode selection tailors response style and depth while maintaining same underlying evidence and rules. All modes obey zero-hallucination principle and cite sources.

Automated Publications Engine

YETO routinely generates structured reports and bulletins on schedules: daily briefs, weekly market updates, monthly economic monitors, quarterly outlooks, annual state-of-economy reports. Each uses predefined templates with dynamic data insertion and AI-drafted narrative summaries.

Daily Brief

Overnight changes, currency movements, key events—internal and premium tier

1

Monthly Monitor

Comprehensive sectoral review—public PDF plus premium detailed version

3

Annual Report

Flagship transparency report—public summary, premium full annexes

4

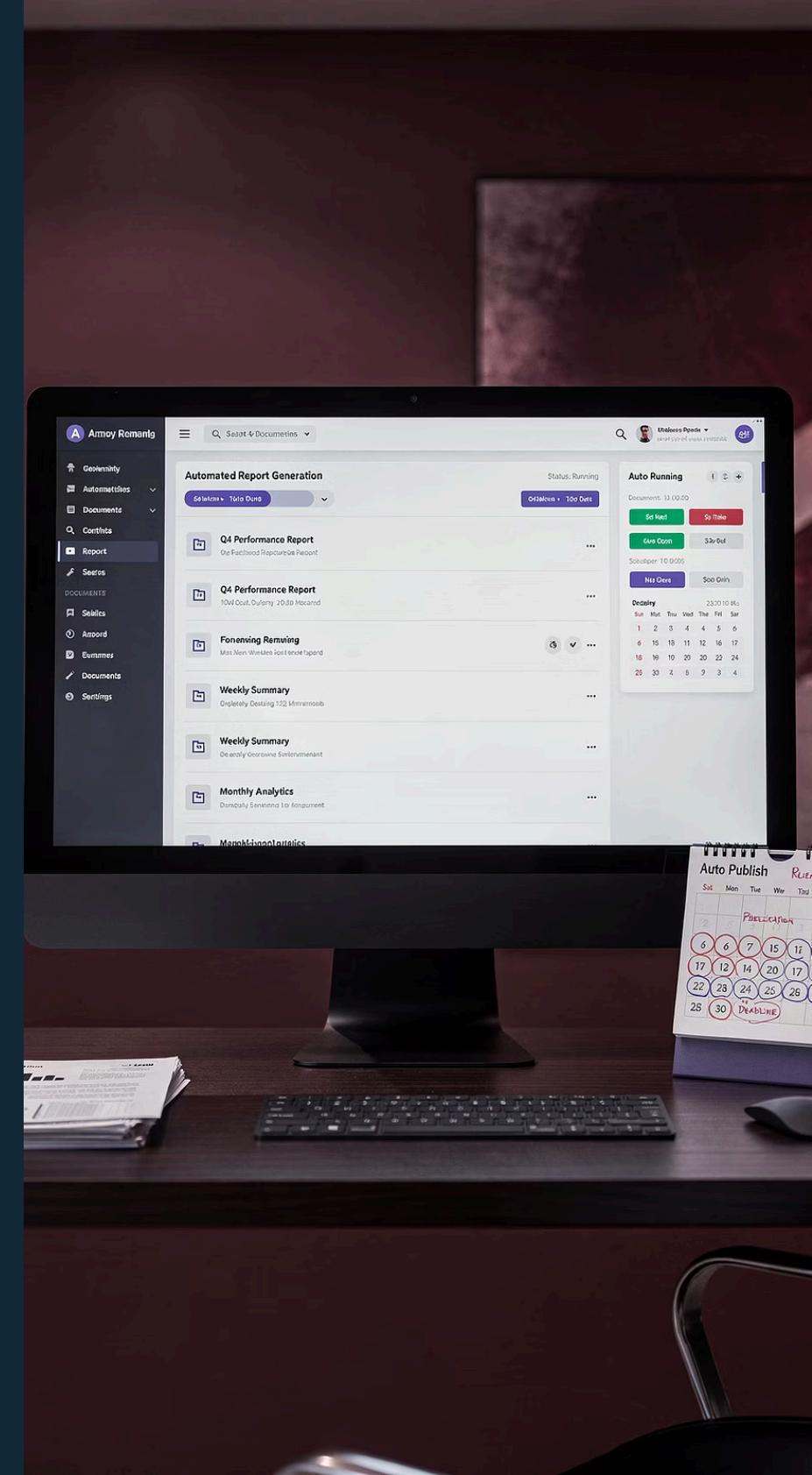
Weekly Update

Market developments, fuel prices, exchange rates—public summary, premium extension

5

Quarterly Outlook

Deep-dive analysis with forecasts and scenarios—premium only





Publication Components and Safeguards

Standard Elements

- Executive summary (5 lines, expandable)
- Charts and tables with source citations
- Discussion of data discrepancies
- Evidence pack appendix
- Data gaps section
- Version and changelog

Safety Gates

If confidence falls below threshold or content could cause panic (bank runs, market manipulation), system queues for admin review before auto-publication. No false alarms published to public without human verification.

User Roles and Access Control



Public

Browse dashboards, read abstracts, limited downloads



Registered Free

Save dashboards, basic alerts, standard exports



Pro Subscriber

Full downloads, advanced analytics, unlimited alerts, premium publications



Institutional

Team workspaces, bulk API, scheduled reports, private datasets



Contributor

Data submission portal, status tracking, restricted dashboards



Administrator

Full system control, user management, ingestion monitoring, audit logs

Subscription Tier Features

Free tier provides transparency fundamentals: browse dashboards, view research library, basic downloads, limited API calls. Pro tier adds saved dashboards, advanced alerts, full data exports, custom reports, premium publications. Institutional tier unlocks team collaboration, bulk API access, scheduled delivery, private datasets, partner integration support.

Payment via Stripe with webhooks for provisioning. Trial periods available. Non-profit discounts offered. Enterprise contracts negotiated individually.

Free	Pro	Enterprise
Noingen	Incunting	Mank Brackn
Ondus	Exelction	Recerfenationg
Skreran	Prededong	Dradiouss
Fcating	Winding	Maken Scanse
Bedk	Sooigoocg	Eawiing Scacks
Cffercozing	Cbcer	Phitime Brack
Rener	Chirg	Secase Toog
Connfimmenion--	Relfatioan	Seass Toog



Easy Yeummme Faie Yemen

Sounes yourmy can fug the yemen
now flyn the ohre oréave to
remeny of Yémenf.

Seeo Knew



Frontend: Homepage Design

Landing page features Yemen contextual imagery with bilingual title "Preserving Yemen's Economic Heritage". Two prominent CTAs: "Explore Dashboard" and "View Latest Research". Four headline KPI cards (GDP Growth, Inflation, Reserves, Labour Force) with sparklines. Contextual imagery tiles link to thematic pages (Trade, Local Economy, Rural Development). Latest Updates feed shows recent publications. Interactive Features Showcase highlights AI Assistant and Scenario Simulator.

Frontend: Economic Dashboard

Multi-functional chart interface with sidebar listing sectors/categories. Filter controls for indicator selection, time range, geographic level, regime toggle. Interactive charts with event annotations and tooltips. Metadata panel on right displays source pack for selected indicator: publisher, dataset, dates, license, confidence, transformations. Export options (PNG, SVG, CSV, Excel, PDF) and "Add to Report" button. Data table below chart shows numerical values with year-on-year changes.



Sector Dashboards



Macroeconomy & Real Economy

GDP, GNI, investment, night-time lights proxy, port throughput



Prices & Cost of Living

CPI, food baskets, fuel prices, spatial price differences by governorate



Monetary & Financial System

Exchange rates (official vs parallel), split-system banking indicators, liquidity



Public Finance

Revenues, expenditures, debt, budget execution, fiscal balance



Trade & External Sector

Imports, exports, trade balance, remittances, port activity



Humanitarian & Social

Food security (IPC), malnutrition, displacement, health, education access

Data Repository Interface

Search and Discovery

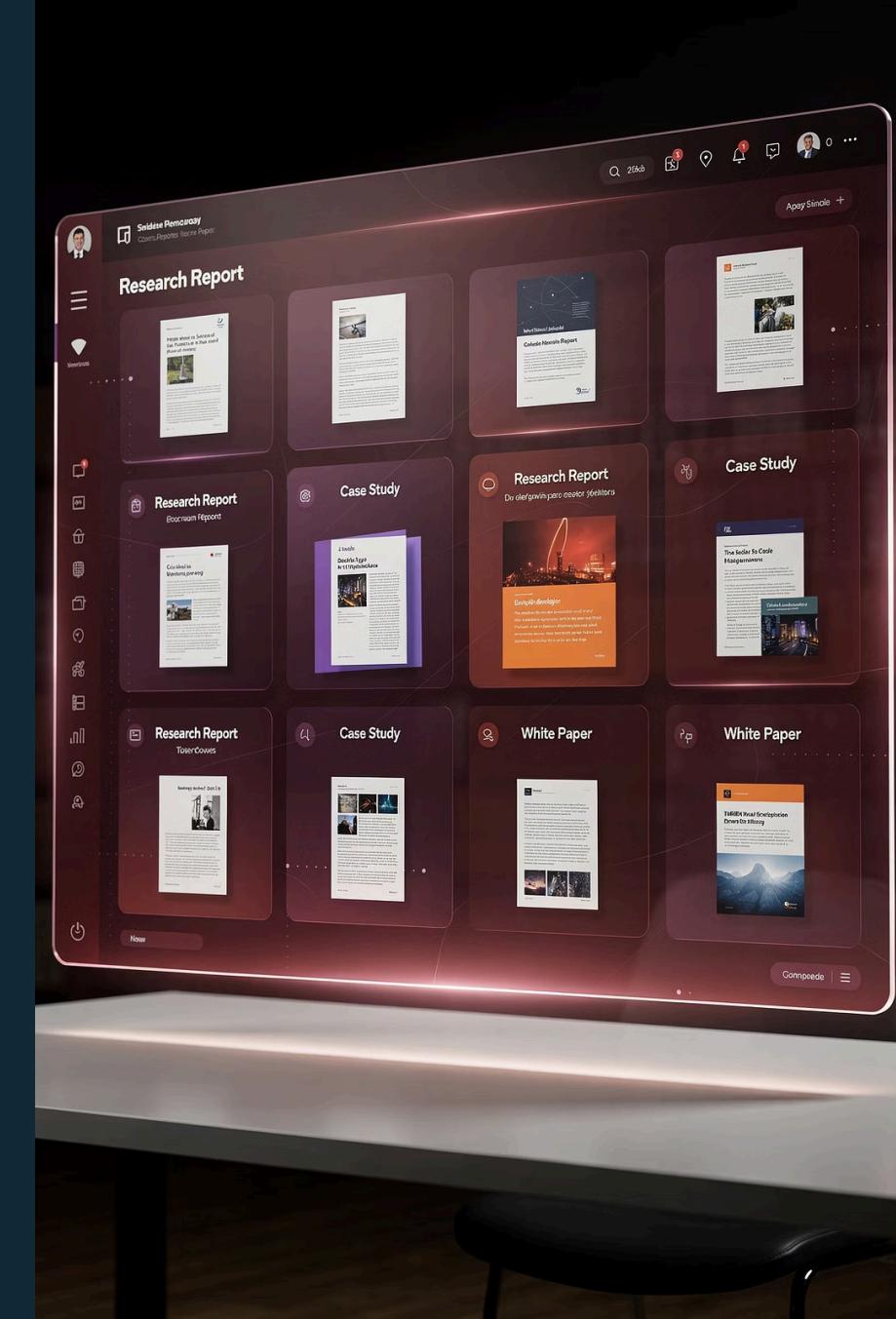
Global search bar with autocomplete. Left panel filters by category, time coverage, geography, data type, quality level. Dataset cards show title, description, verification badges, update status, download counts. Sorting by relevance, date added, popularity.

Dataset Detail View

Full metadata display: source, definition, coverage, frequency, license. Preview chart or table. API endpoint example with cURL command. Download options (CSV, Excel, JSON). Version history for datasets with revisions. Comments or discussion threads if implemented.

Research Library and Reports Hub

Category filters (Economic Policy, Financial Sector, Trade, Poverty, Governance, Conflict Economics). Featured Research highlights at top. Research cards display title, authors, date, abstract snippet, source tags. "Read Full Report" opens detail page with metadata, inline PDF viewer, related datasets, evidence pack, download link. Bilingual content toggle. Full-text search across OCR-extracted documents.



Interactive Timeline of Events



2014: Conflict Escalation

Fighting intensifies; markets destabilize



2016: Currency Split

Monetary division triggers inflation spike



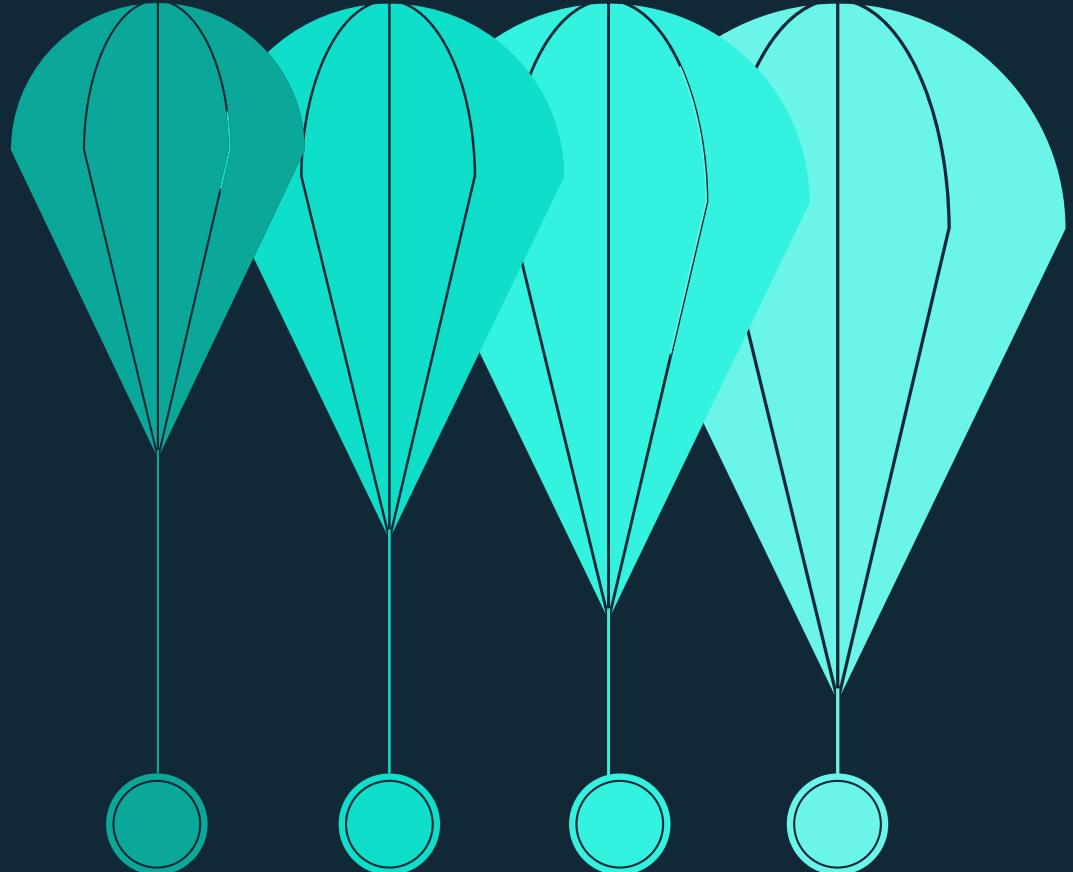
2018: Humanitarian Crisis

Displacement rises; aid flows strain economy



2021–Present: Policy Changes

Reforms alter trade and fiscal balances



Chronological view of significant events with cards showing icon, title, date. Colour-coded by category (political, economic policy, conflict, humanitarian). Filters by actor, location, theme. Clicking event expands to show description, sources, linked indicators with before/after comparison. "Neutralise" toggle for counterfactual scenarios—exclude events and see alternative trend lines via scenario simulator.

Glossary: Bilingual Economic Terms

Navigation

Alphabetical index with letter jump links.
Category filters (macroeconomic,
humanitarian, financial terms). Search bar
with instant results.

Term Detail

English and Arabic terms displayed. Definition in both languages. Yemen-specific context explaining local
nuances. Related indicators with links. Related terms as clickable tags. Optional embedded chart
illustrating concept. References to external glossaries or sources.

Methodology and Transparency Page

Data Collection

Overview of sources, ingestion frequency, pipeline architecture.
Source Registry access for exploring catalogue.

Quality Assurance

Explanation of verification scoring
(A/B/C/D), outlier handling,
contradiction management. QA
dashboard widgets showing platform
health.

Transparency Pillars

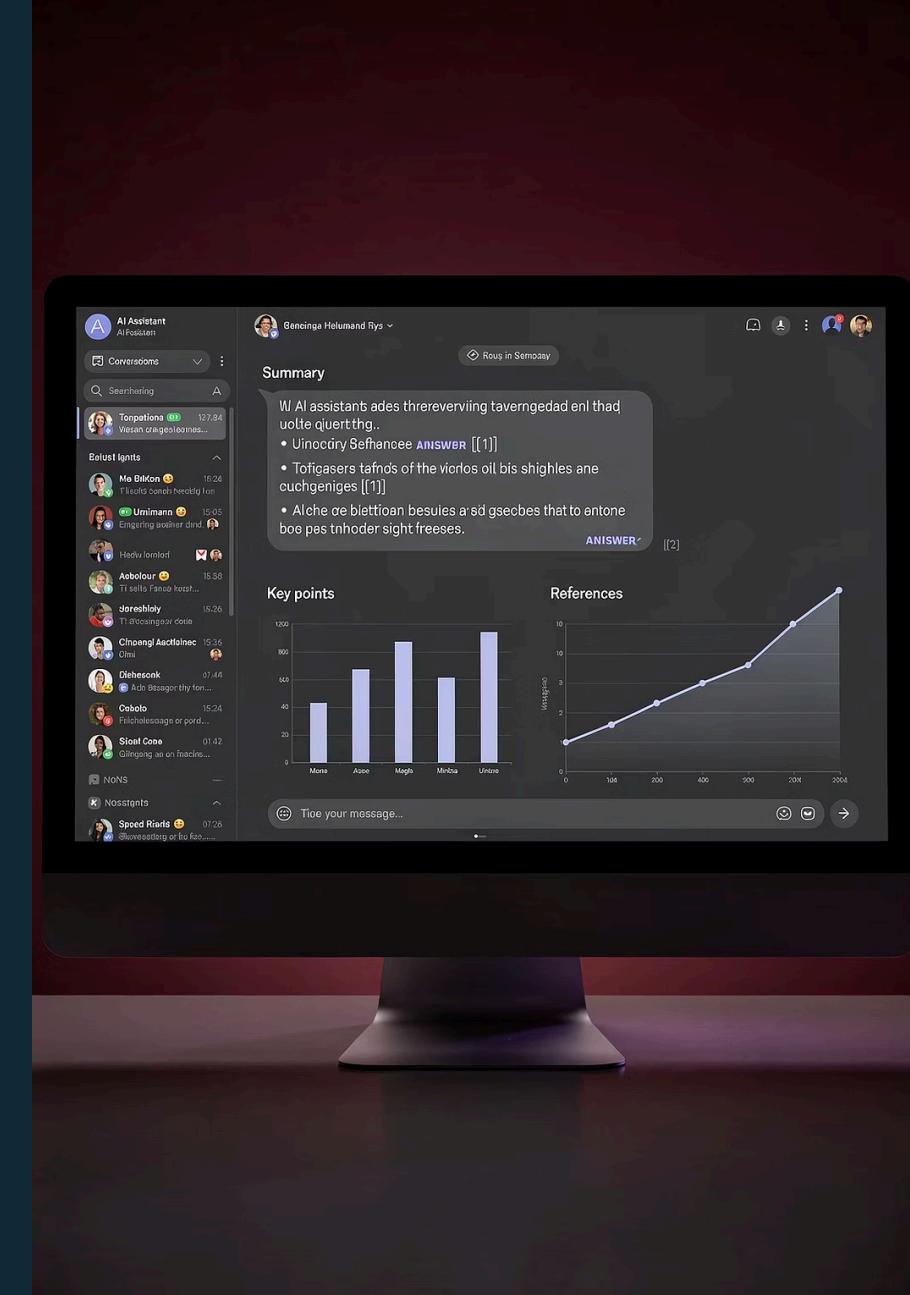
Provenance tracking, public corrections policy, versioning system. "Where does this number come from?" philosophy explained.

Expandable sections with technical details. Coverage completeness stats. "Contact Data Team" CTA for questions or data contributions. Link to changelog for significant methodology updates.



AI Assistant: Chat Interface

Split-pane layout: left shows conversation history, middle displays chat thread with user questions and structured AI responses (9-part format), right panel offers Agent Mode selector and suggested questions. Input box at bottom for queries. Export and Share buttons. Language toggle (EN/AR). "Show Evidence Pack" button reveals underlying sources for each answer. Evidence includes inline charts with confidence badges.



Custom Report Builder Wizard



Select Data

Choose indicators or datasets from catalogue



Choose Visualizations

Pick chart types for each data element



Customize Layout

Drag-arrange charts, add text blocks, set title and options



Export

Generate PDF, Word, PowerPoint, Excel with embedded citations

Report Settings panel: title, author, date range, colour scheme, language (English/Arabic/bilingual). Automatic source citations aggregated at end or per chart. Preview before final generation.



Scenario Simulator

Simulation Setup

Select economic domain (macro, fiscal, humanitarian). Adjust assumptions via sliders: oil price change, port throughput increase, aid flow adjustments, remittance levels. Choose preset scenarios (Best Case, Worst Case) or custom.

Results and Analysis

Comparison charts: baseline vs scenario outcomes. Narrative summary generated by AI. Sensitivity analysis showing how results vary with assumption changes. Explicit labelling of assumptions not evidence-backed. Export scenario as report.

The screenshot displays the Admin Portal interface. On the left, a sidebar lists navigation options: Sources, Retries, Review Queue (selected), Readings, Closeby, Sendings, and Helpcenter. The main area is divided into two sections: 'Source Registry' and 'Review Queue'.
Source Registry: A table with columns: Name, Type, Credibility Rating, Update Cadence, License, Actions. It lists sources like '100H 0400 LS2202', '100H 0403 LS2202', etc., with actions 'Edit' and 'Disable'.
Review Queue: A table with columns: Name, Type, Credibility Rating, Update Cadence, License, Actions. It lists sources like '100H N025', '100H 89MT', etc., with automated metadata suggestions (e.g., 'SourceName: 26211', 'Score: 43.20.10') and 'Approve' or 'Reject' buttons.
Ingestion Monitor: A section showing recent job statuses, logs, and next scheduled runs, with manual trigger options.

Name	Type	Credibility Rating	Update Cadence	License	Actions
100H 0400 LS2202	Source Name	5870	50.50.10	22023	11 080531
100H 0403 LS2202	Source Name	5945	13.20.15	22004	11 620203
100H N025	Source Name	26211	43.20.10	20034	11 090272
100H 89MT	Source Name	26000	50.20.13	20204	15 031203
100H 80MT	Source Name	5607	70.20.20	20034	11 622285
100H 0400 LS2202	Source Name	56810	50.30.15	52063	17 637255
100H 0400 LS2202	Source Name	56817	20.20.15	52024	11 630105
100H 0400 LS2202	Source Name	26745	20.20.40	20063	11 042296

Admin Portal: Source Management

Source Registry table: name, type, credibility rating, update cadence, license, actions (edit, disable). Review Queue lists newly discovered sources with automated metadata suggestions and approve/reject buttons. Ingestion Monitor displays recent job statuses, logs, next scheduled runs, manual trigger options. Integration with Source Discovery Engine for weekly scans.

Admin Portal: Quality Dashboards

Outliers Flagged

List of anomalous data points awaiting review with series, date, value, reason flagged

Contradictions Detected

Conflicts between sources with side-by-side values, resolution status

Coverage Heatmap

Matrix of indicators vs years showing data completeness percentage

Gap Tickets

Open tickets for missing data with description, priority, suggested sources, status

Model Monitoring

Forecast error trends, drift alerts for ML models, backtesting results

Audit Logs

Security events, admin actions, user permission changes with timestamps





Admin Portal: User and Subscription Management

Search users by email or organisation.

View roles and subscription tiers.

Actions: upgrade/downgrade tier, assign roles (contributor, admin), reset passwords, revoke access. Integration with Cognito for identity management and Stripe for payment status. Trial management: see who's on trial, extend or convert to paid.

Publication Approvals

Queue of auto-generated reports flagged for review (low confidence or sensitive content). Preview content, edit if needed, approve for publication or reject with feedback. Once approved, publication posts to site and triggers email distribution.

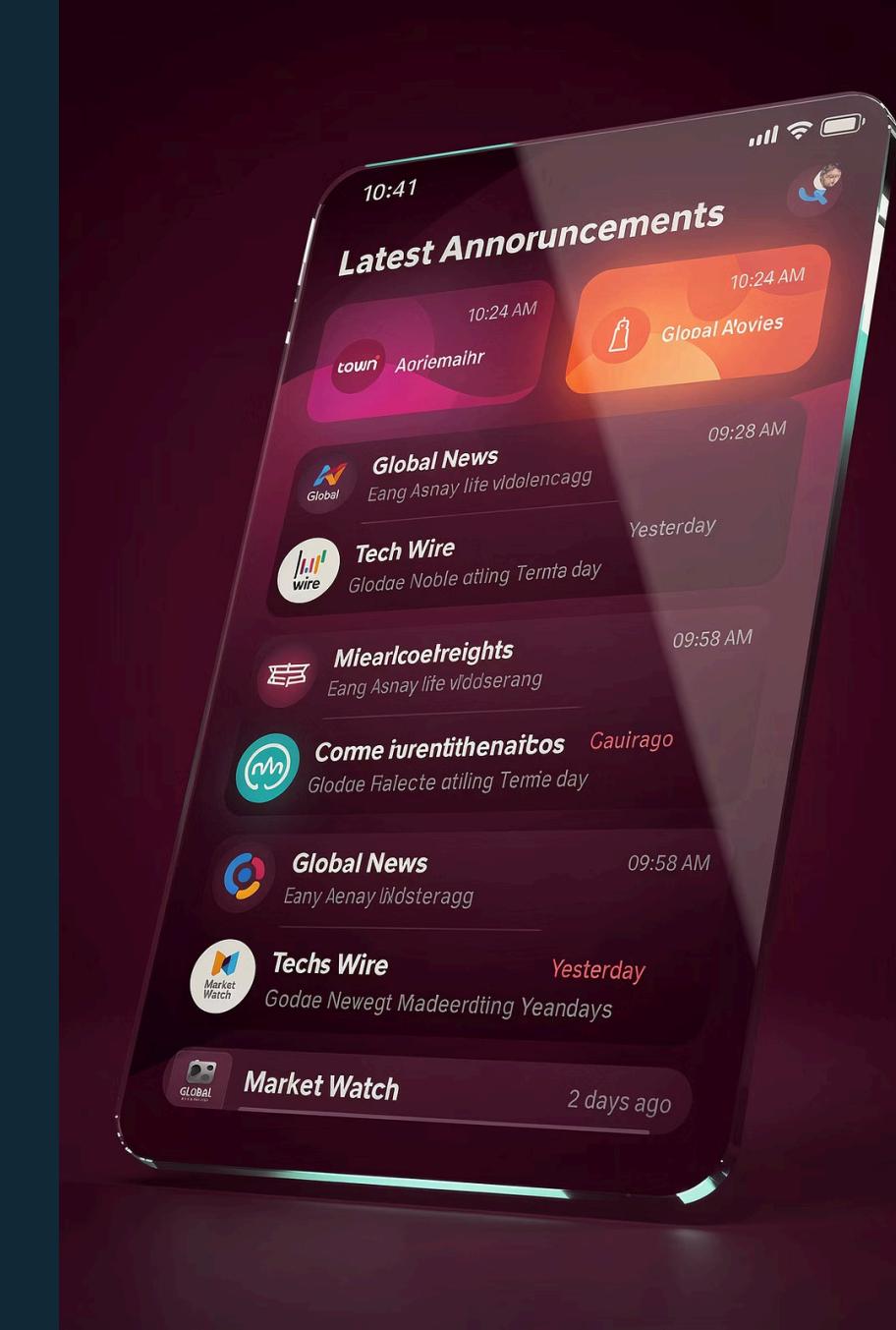
Contributor Portal: Data Submission

Upload interface: drag-and-drop CSV/Excel or manual entry. Metadata form: indicator name, unit, time coverage, geographic scope, source authority. Submission Status dashboard shows submitted datasets with pending/approved/needs clarification tags. Communication thread with YETO team for questions or feedback on submissions. Restricted dashboards visible only to contributor's organisation showing their data integrated with broader context.



Latest Updates Feed

Chronological grid of update cards: YETO announcements, new data additions, insights posts, external news aggregated from official sources (World Bank press releases, IMF statements). Each card includes image, source icon, title, excerpt, date/time, "Read more" link. Filters by type (insight, news, data update). Automated ingestion pipeline populates feed alongside internal posts.





Security Architecture: Defence in Depth

Application Layer

1

Input validation, HTTPS only, JWT tokens, role-based access

Network Layer

2

WAF filtering, rate limiting, CloudFront DDoS protection, ALB security groups

Data Layer

3

Encryption at rest (RDS, S3, OpenSearch), TLS in transit, KMS-managed keys

Identity Layer

4

Cognito user pools, MFA optional, IAM least privilege, secrets in Secrets Manager

Monitoring Layer

5

CloudWatch logs, audit trails, CloudTrail API monitoring, automated alerting

Security: Secrets and Encryption

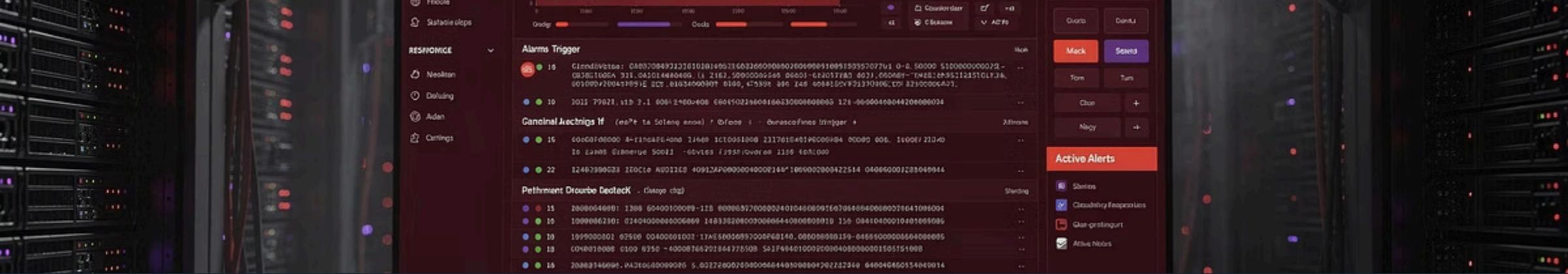
Secrets Management

All credentials stored in AWS Secrets Manager: database passwords, API keys for external sources, encryption keys. No hard-coded secrets in code repositories. IAM roles grant services access only to needed secrets. Automatic rotation for database credentials.

Encryption Standards

RDS PostgreSQL: KMS-managed transparent encryption. S3 buckets: default AES-256 or KMS encryption. OpenSearch: encrypted indices. CloudFront and ALB: TLS 1.2+ with ACM certificates. Internal service calls: TLS or VPC security.



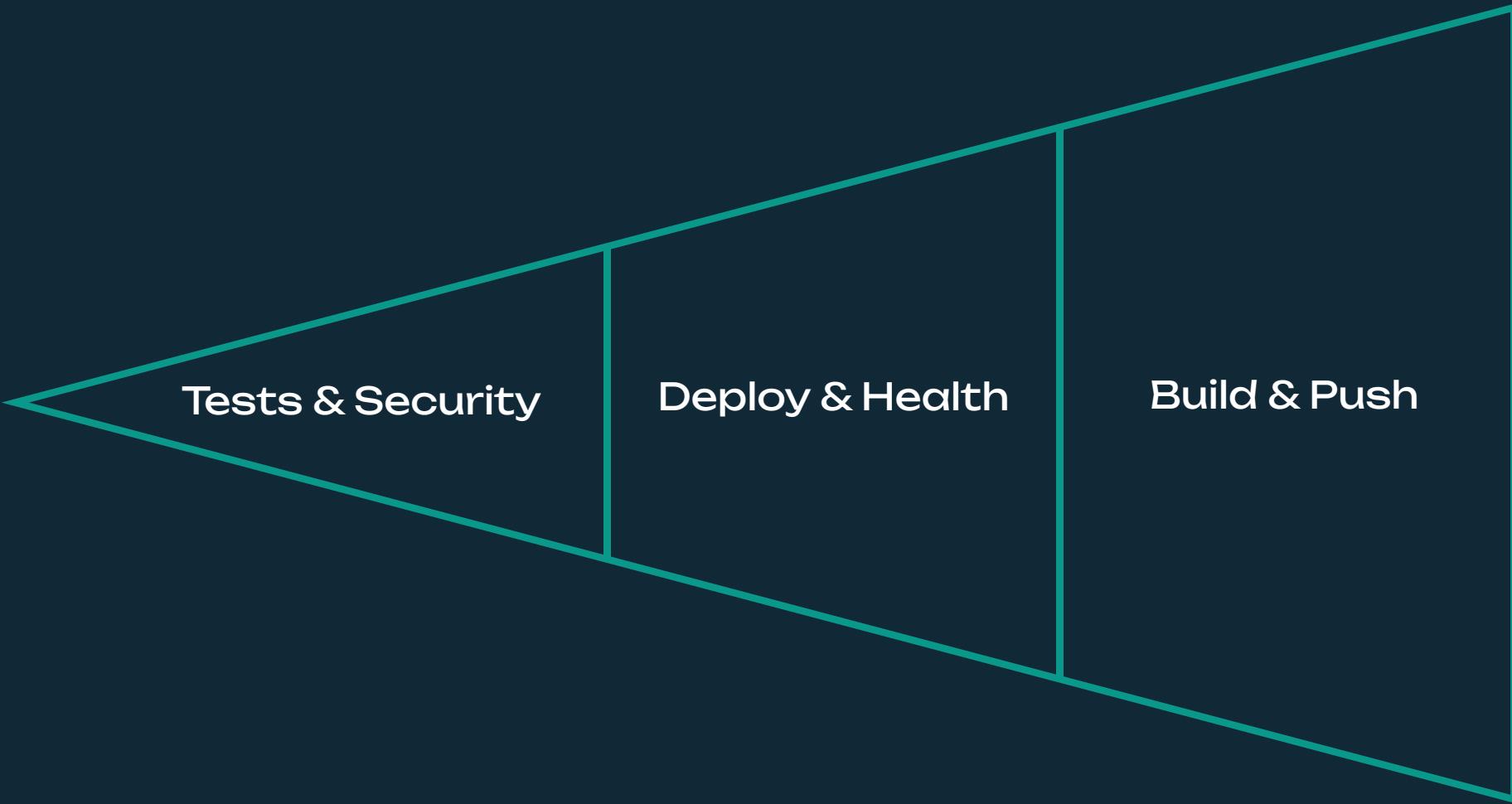


Security: Monitoring and Incident Response

CloudWatch aggregates logs from all services. Automated alarms trigger for anomalies: sudden traffic spike, multiple failed login attempts, unusual API patterns. CloudTrail logs all AWS console and API activity for audit. Application audit logs record admin actions in database visible via admin portal.

Incident Response Runbook provides step-by-step procedures for common scenarios: data breach, DDoS attack, site downtime, unauthorized access. SNS/email alerts notify devops team immediately. Backup and recovery processes documented and tested quarterly.

DevOps: CI/CD Pipeline Architecture



GitHub Actions orchestrates pipeline. Main branch protected—requires pull request with passing checks. Separate staging and production environments. Infrastructure-as-code via Terraform ensures reproducible deployments. Rollbacks possible via previous container images and database snapshots.

DevOps: Testing Strategy

Unit Tests

Data transforms, calculations, connector logic—Jest for JS, PyTest for Python

Integration Tests

API endpoints, database operations—test environment with sample data

E2E Tests

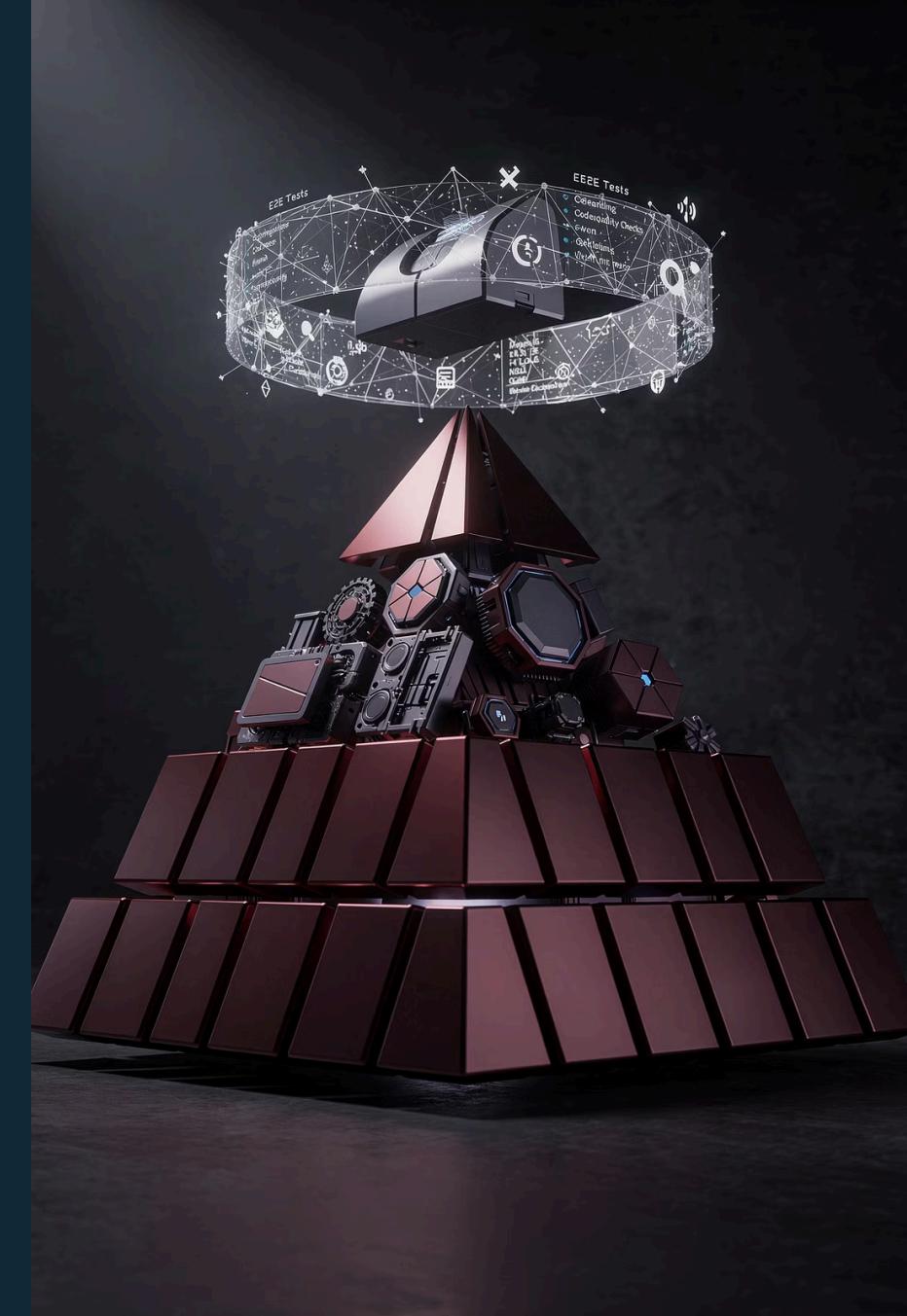
Critical user flows—Playwright or Cypress simulating browser interactions

Security Scans

SAST, dependency vulnerabilities—automated in CI pipeline

Nightly QA

Data integrity checks, coverage scorecard validation, fake data detection





Infrastructure: AWS Deployment Stack

Compute

- AWS App Runner or ECS Fargate for containerised services
 - Lambda for event-driven ingestion jobs
 - EventBridge for scheduled tasks
 - Step Functions for multi-step workflows

Storage & Data

- RDS PostgreSQL (encrypted, automated backups, read replicas)
 - S3 (raw/processed data, documents, versioning enabled)
 - OpenSearch Service (full-text search, semantic vectors)
 - Secrets Manager for credentials

CloudFront CDN serves static assets. Route53 manages DNS. ACM provides SSL certificates. VPC with security groups isolates resources. CloudWatch monitors all services. Infrastructure defined in Terraform for repeatability.

Documentation Deliverables



Administrator Manual

Portal usage, source management, user administration, troubleshooting, backup procedures



Data Governance Manual

Lifecycle policies, confidence assignment, contradiction resolution, versioning protocols



API Manual

Endpoint documentation, example requests/responses, rate limits, SDK usage, terms of use



Contributor Manual

Data formatting templates, submission portal guide, verification process, data sharing agreements



Subscriber Manual

Premium features guide, alert setup, workspace usage, report builder, subscription management



Model Cards Library

AI/ML model documentation—training data, algorithms, accuracy metrics, limitations



Recovery Runbook

Disaster recovery procedures, database restoration, security breach response, RTO/RPO targets

All documentation versioned in Git, bilingual for user-facing materials, auto-updated when system changes.

Stakeholder Engagement Strategy

Comprehensive Stakeholder Map identifies donors, IFIs, UN agencies, Yemeni institutions, private sector, academia, NGOs. For each: tailored engagement plan addressing their interests and what YETO needs from them (data access, technical feedback, funding).

Outreach materials include one-pagers (non-technical and technical), methodology documents, stakeholder map, email templates, pitch deck, budget plan, job descriptions. Bilateral meetings, launch events, webinars demonstrate platform value and build partnerships.



Key Stakeholder Categories



International Financial Institutions

World Bank, IMF, Arab Monetary Fund—data sources and users, potential co-funding partners



UN Agencies

UNDP, WFP, OCHA, UNICEF, FAO—humanitarian data, funding flows, joint analytical partnerships



Yemeni Institutions

Central Banks (Aden & Sana'a), ministries, CSO—primary data sources, beneficiaries of analysis



Donors

EU, USAID, Gulf states—governance focus, accountability tools, funding for platform sustainability



Research & Think Tanks

Sana'a Center, Rethinking Yemen's Economy—analytical content, data users, co-publication opportunities



Private Sector

Banks, chambers of commerce, business associations—market data sources, compliance monitoring users

Go-Live Quality Checklist

No Fake Statistics

Verify no placeholder or dummy data remains visible. Test script crawls site confirming all KPIs show real values or explicit "Not available yet".

Source Pack Completeness

Every chart and figure has accessible metadata: publisher, dataset, dates, license, confidence, transformations.

Contradictions Displayed

Conflicting data points visible via disagreement mode or notes. No hidden suppression of discrepancies.

Arabic UX Verified

Native Arabic speakers test full functionality: proper RTL, correct glossary terms, readable content mirror of English.

Exports Include Citations

All download formats (CSV, PDF, images) embed source citations and retrieval dates automatically.

Ingestion Resilience

Simulate failures and restarts. Confirm idempotency, retry logic, proper error logging without data loss.

Backups Configured

RDS automated backups tested, S3 versioning enabled, OpenSearch snapshots scheduled. Recovery runbook validated.

Security Hardened

IAM least privilege verified, encryption confirmed, WAF active, audit logs capturing events.

CI Green

Main branch tests passing 100%. No flaky tests or known failures.

Design Consistency

All pages match design system: navy/green/gold palette, clean cards, consistent typography and spacing.





Coverage and Data Population

Deployment Gate

Minimum 70% coverage of core indicators (2010–present) or explicit gap documentation for missing areas. Coverage scorecard validates data completeness across sectors and time periods.

Initial Backfill Complete

Major datasets ingested: macroeconomic series from World Bank/IMF, humanitarian indicators from UN agencies, trade statistics, aid flows from OCHA FTS/IATI, conflict events from ACLED/UCDP, research library populated with key reports. Gap tickets created for known missing data with explanations visible to users.



Sustainability: 5-Year Budget and Staffing

Financial plan includes CAPEX (initial development, cloud setup) and OPEX (hosting costs, data licenses, salaries, maintenance). Three scenarios: lean (minimal team, basic features), base (full team, all features), ambitious (expansion with regional coverage). Headcount by role: Chief Economist, Data Engineering Lead, Backend/Frontend/MLOps Leads, Research Director, QA Lead, Product Manager, UX Lead, Partnerships Lead, Security Lead, Editor-in-Chief.

Funding sources: donor grants, institutional subscriptions, foundation support. Break-even projections and reserve for risk management. Cashflow projection with quarterly milestones.



Team Structure and Key Roles

	Chief Economist Analytical vision, methodology design, model oversight, stakeholder engagement
	Data Engineering Lead Pipeline architecture, connector development, ETL optimization, data quality
	Backend Lead API design, authentication, ingestion orchestration, performance optimization
	Frontend Lead UI implementation, dashboard interactivity, responsive design, bilingual UX
	MLOps/AI Lead AI assistant development, RAG system, model training, drift monitoring
	Research Director Content curation, analysis quality, publication oversight, expert review
	QA/Verification Lead Data validation, outlier review, contradiction resolution, coverage monitoring
	Partnerships Lead Stakeholder outreach, MoU negotiation, contributor onboarding, donor relations

Outreach Materials Suite

One-Pager (External)

Non-technical overview: mission, unique features, bilingual brochure for general audiences

Technical One-Pager

Architecture, security posture, API design—for IT departments and technical reviewers

Methodology Document

Plain version for trust-building, technical deep dive for rigorous scrutiny

Stakeholder Map

Excel spreadsheet and PDF summary identifying partners, influence, engagement plans

Email Templates

Tailored outreach for each stakeholder type: donors, IFIs, UN agencies, local institutions

Pitch Deck

10–14 slides: problem, solution, governance, roadmap, budget, ask (data/funding/feedback)





YETO's Transformative Impact

For Policymakers

Evidence-based decision support. Real-time economic monitoring.
Scenario planning for policy interventions. Split-system clarity
preventing conflation of incompatible data.

For Donors

Accountability tracking of aid flows and project outcomes. Transparent
needs assessments. Coordination tool reducing duplication and
information gaps.

For Researchers

Comprehensive dataset catalogue with full provenance. Citation-ready
exports. Access to documents and analysis previously scattered or
inaccessible.

For Yemeni Public

Arabic-first interface making economic data accessible locally.
Transparency into governance and resource flows. Countering
misinformation with verified evidence.



Platform Readiness: The Stop Condition

YETO is considered "done" and ready for launch when all quality gates are satisfied and the platform operates autonomously:

100... 70%+ 24/7

Data Integrity

No fake statistics visible.
Every number has source
pack. Contradictions
displayed transparently.

Coverage Minimum

Core indicators populated
2010–present or gaps
explicitly documented with
tickets.

Continuous Operation

Auto-updating data
pipelines, scheduled
publications, resilient via
backups, manageable via
admin console without
developer intervention.

The live site at yeto.causewaygrp.com is populated with real content, fully functional in English and Arabic, and meets all twelve absolute rules. The platform is now self-sustaining, delivering decision-grade intelligence to stakeholders across Yemen and the international community.



A New Standard for Economic Transparency

YETO represents a paradigm shift in how fragmented, conflict-affected economies are understood and monitored. By refusing to fabricate, hiding nothing, and treating division as reality rather than aberration, it establishes a new standard for evidence-based intelligence platforms.

This is not a data portal. It is a constitutional-level commitment to truth in the service of better decisions. Every architectural choice—split-system enforcement, provenance tracking, contradiction display, Arabic-first design—reflects a refusal to compromise on transparency for convenience.

The platform is now ready to serve its mission: preserving Yemen's economic heritage and resilience through crisis, one verified data point at a time.