PHREVO PLATFORM

Technical Architecture Annex

Systems, Data Infrastructure & Machine Learning Methods · 2026

This annex provides a comprehensive technical reference for the PhREVO platform, covering the architecture of all major subsystems and the full catalogue of machine learning algorithms deployed across the donor, project, verification, and exchange layers. It is intended as a companion document to the PhREVO Master Synthesis 2026 and the Public Manual.

PART I

Technical Architecture

PhREVO is structured as five interconnected subsystems that form a closed-loop pipeline: data about donors and projects enters the platform, is verified and scored, then converted into investable impact portfolios traded on the Impact Exchange. Each subsystem is described below with its inputs, outputs, data infrastructure, and key technologies.

1.1 Platform Overview

The platform operates on three cloud environments (AWS, GCP, and Azure integration services) with a polyglot persistence layer, an AI/ML engine, and full-stack mobile and web frontends. Security and compliance are addressed at every tier — from encrypted storage to identity federation and IoT endpoint governance.

Subsystem	Primary Function	Key Outputs
Donor Data System	Profile and match altruistic investors to projects	SDG-matched donor profiles, portfolio preferences
External Data System	Ingest and classify global project data via neural networks	SDG-scored project candidates
Project Management System	Track on-the-ground delivery via IoT and field tools	Volunteer scores, sensor data, field stories
Project Verification System	Apply three-score impact model to projects	Total Project Value (Score 1·2·3)
Stock Exchange Trading System	Convert verified projects into tradeable impact portfolios	Portfolio prices, investment instruments
Altruistic Investor System	Personalised investor interface and portfolio management	Investment decisions, PhREVO coin transactions

1.2 Cloud Infrastructure

PhREVO runs a multi-cloud strategy to achieve geographic resilience, regulatory compliance, and specialised service coverage.

Provider	Services Used	Rationale
Amazon Web Services (AWS)	DynamoDB, EC2, Lambda, S3, Kinesis, SageMaker	Core compute and managed NoSQL for high-throughput event streams

Google Cloud Platform (GCP)	Bigtable, BigQuery, Vertex AI, Pub/Sub	Wide-column analytics and large-scale ML training workloads
Azure Integration Layer	Service Bus, Event Grid, API Management, Logic Apps, PowerBI, Stream Analytics	Enterprise integration, workflow orchestration, and executive dashboards

1.3 Database Architecture

A polyglot persistence model matches each data shape to the most appropriate storage engine, avoiding the impedance mismatch of a single-database approach.

Category	Type	Engine	Primary Use Case
NoSQL — Document		**MongoDB**	Project records, donor profiles, unstructured field data
NoSQL — Key-Value		**Redis**	Session cache, real-time leaderboards, ephemeral scoring
NoSQL — Wide-column		**Cassandra / Bigtable**	Time-series IoT sensor telemetry at scale
NoSQL — Graph		**Neo4J**	User-to-project relationship mapping, SDG adjacency graphs
Relational (SQL)		**MySQL / MariaDB**	Transactional data, investor accounts, financial ledger
Relational (SQL)		**PostgreSQL**	Geospatial queries, advanced analytics, data modelling
Real-time NoSQL		**Firebase Realtime DB**	Mobile sync, live volunteer scoring updates
Search		**Elasticsearch**	Full-text project search, log aggregation, SDG keyword matching

1.4 Security Architecture

Security is enforced across six layers, aligned with GDPR and PCI-DSS requirements.

Security Layer	Controls & Technologies
Data Protection	AES-256 encryption at rest and in transit, data loss prevention, blockchain-anchored audit trails, key vault with HSM, data access governance
Identity & Access Management	Azure Active Directory, AD Domain Services, AD B2C, managed identities, application registrations, RBAC

Application Security	Web application firewall, component vulnerability scanning, OWASP Top 10 mitigations, SAST/DAST pipelines
Cloud & Container Security	Container hardening, cloud workspace compliance, CIS benchmark controls
Network Security	Firewall, TLS/SSL everywhere, DDoS mitigation, remote access via zero-trust VPN
Endpoint & IoT Security	Device SDK authentication, Arduino/Raspberry Pi gateway controls, mobile device management for iOS and Android field tools
Security Operations	24/7 monitoring, vulnerability management, threat detection and analysis, incident response playbooks, SOAR automation
Compliance	GDPR (data subject rights, DPO designation, breach notification), PCI-DSS (PhREVO Coin wallet transactions)

1.5 Development Stack

Backend

- Python — data engineering, ETL, ML pipelines (NumPy, SciPy, Django, Flask)
- Java / Spring ecosystem — Spring Boot, Spring Security, Spring Cloud, Spring Data, Jakarta EE, Hibernate ORM
- Node.js / Express.js — lightweight API services and event-driven microservices
- PHP (Laravel, Symfony) — NGO-facing web portals and CMS integrations
- C# / .NET Core — Microsoft-stack services and Azure Logic App integrations

Frontend

- Web: React, Vue.js, Angular
- iOS: Swift / Objective-C (Xcode, CodeRunner); apps for iPhone, iPad, Apple Watch, Apple TV
- Android: Kotlin / Java (Android Studio, Eclipse)
- Cross-platform: JavaScript / C# (shared business logic)

IoT & Edge

- Arduino — field sensor integration and data capture
- Raspberry Pi — edge gateway for remote project sites
- Device Gateway + Device SDK — authentication, telemetry ingestion, OTA updates

Integration & Messaging (Azure)

- Service Bus & Event Grid — decoupled event-driven architecture

- API Management — unified API gateway for third-party data providers
- Logic Apps & Integration Accounts — low-code workflow automation

PART II

Machine Learning & Algorithms

Each PhREVO subsystem deploys a tailored set of machine learning algorithms matched to the specific data shapes and decision tasks it handles. The table below catalogues every algorithm by subsystem, task, and rationale.

2.1 Donor Data System

The Donor Data System builds rich, multi-dimensional profiles of altruistic investors — individuals and companies — and matches them to projects aligned with their values, risk profile, and SDG priorities.

Task	Algorithm(s)	Rationale
Information gathering & profiling	Self-trained Naive Bayes Classifier / NLP (Text Mining)	Probabilistic text classification of survey responses, digital footprint, and social signals into investor interest categories
Areas of influence mapping	Random Forest	Ensemble method handles mixed data types (demographics, behavioural, geographic) with built-in feature importance for explainability
SDG matching	Fuzzy C-Means Clustering	Soft clustering allows donors to belong partially to multiple SDG affinities, reflecting real-world overlap in values
Connecting with other users	Hierarchical Clustering / Multivariate Regression	Identifies peer communities and predicts affinity scores between investors for social referral and co-investment features
User activity prediction	Artificial Neural Networks / Collaborative Filtering	ANN captures non-linear behavioural patterns; collaborative filtering surfaces projects similar donors have engaged with
Portfolio interaction optimisation	Logistic Regression / Deep Learning	Logistic regression for binary engagement decisions (invest / skip); deep learning for multi-step portfolio behaviour modelling

2.2 External Data System

The External Data System ingests project submissions from NGOs and external data providers and runs a multi-stage ML pipeline to classify, cluster, and score them before they enter the verification queue.

Task	Algorithm(s)	Rationale
Data gathering & structure inference	Gaussian Mixture Model / Expectation-Maximisation (EM)	Probabilistic approach for discovering latent project clusters when ground-truth labels are absent in new data sources
Project data classification	K-Nearest Neighbors (KNN) / Support Vector Machine (SVM)	KNN for fast prototype-based classification of SDG categories; SVM for high-dimensional text feature spaces
SDG application (multi-label scoring)	Multivariate Regression	Each project receives a continuous alignment score across all 17 SDGs simultaneously, enabling partial and overlapping SDG attribution
Multiple target identification	DBSCAN (Density-Based Spatial Clustering)	Detects clusters of project beneficiaries at arbitrary density without pre-specifying cluster count; identifies outlier projects
Blockchain consensus	Byzantine Fault-Tolerant Consensus	Ensures data integrity across distributed verification nodes; any node's data is cross-validated before entering the ledger

2.3 Project Verification System — Three-Score Model

The Project Verification System is the methodological core of PhREVO. Every project must be measured by specific, unique, and standardised indicators against a pre-defined baseline. Results are composed into three hierarchical scores that together produce the Total Project Value.

Score 1 — Influence Areas

Task	Algorithm(s)	Rationale
Project information	Self-trained Naive Bayes / NLP	Classifies project narrative,

classification	Text Mining	objectives, and descriptions into standardised impact categories using domain-specific training corpora
SDG classification	K-Nearest Neighbors (KNN)	Instance-based learning assigns each project to the most similar verified SDG category based on feature proximity
Areas of influence mapping	Random Forest	Identifies which geographic, demographic, and thematic influence zones a project activates
Qualification tools	SVM / Logistic Regression	Binary and multi-class classification for project eligibility gates and certification tiers
Thematics identification	A-priori Algorithm (Association Rule Mining)	Discovers co-occurring thematic patterns across projects, enabling portfolio-level thematic diversification

Score 2 — Benchmark

Task	Algorithm(s)	Rationale
Number of individuals estimation	Lasso Regression	Regularised regression selects the most predictive features of beneficiary reach while controlling for collinearity in social data
Duration estimation	Model-based value estimation	Bayesian priors from historical project durations are updated with project-specific features to produce duration probability distributions
Financing structure	Decision Tree / Linear Regression	Decision tree identifies financing pathway (grant, debt, equity, blended); linear regression estimates resource requirements
Peer deviation scoring	Clustering + Deviation % analysis	Projects are clustered by type and geography; deviation from cluster centroid (expressed as %) becomes the benchmark score

Score 3 — Impact

Impact scoring applies standardised indicators with pre-defined baselines. Three performance thresholds are tracked for each indicator: Lower Limit, Optimal, and Exceptional. The composite Score 3 aggregates indicator performance weighted by SDG criticality and economy data.

Component	Definition
Criticality index	Weighted composite of critical SDG alignment, critical indicators breached, historical performance trajectory, and macroeconomic context data
Category & targets	Direct and indirect beneficiary categories, quantification tools applied, and IRS+ methodology alignment
Impact thresholds	Lower Limit / Optimal / Exceptional — three-band performance assessment against the project's own pre-registered baseline

2.4 Project Management System

The Project Management System uses lightweight algorithms to support on-the-ground data collection, volunteer scoring, and real-time IoT telemetry processing.

Task	Algorithm(s)	Rationale
Data gathering & aggregation	Gaussian Mixture Model / EM Algorithm	Handles mixed, partially labelled field data with missing values common in low-connectivity environments
Data classification	KNN / SVM	Fast, low-compute classification suitable for edge inference on Raspberry Pi field devices
SDG application	Multivariate Regression	Maps field observations to SDG indicator movements in real time
Multiple target tracking	DBSCAN	Identifies beneficiary clusters and detects anomalies in sensor streams
Blockchain verification	Consensus algorithm	Field data submissions are hashed and anchored on-chain for tamper-proof audit trails

2.5 Stock Exchange Trading System

The Stock Exchange Trading System converts verified project scores into tradeable instruments. It uses statistical and financial modelling to manage portfolio value, detect behavioural anomalies, and match supply and demand for impact assets.

- Portfolio Combination System — combinatorial optimisation selects project bundles that maximise impact-adjusted return within donor-specified SDG constraints.
- Project Behaviour Management — time-series analysis (ARIMA, rolling statistics) monitors project performance against historical price curves (JAN–JUL reference windows).
- Stock value estimation — model-based valuations combine Score 1·2·3 outputs with macroeconomic and SDG criticality weights to produce a continuous project price.
- Statistics & Probability engine — Monte Carlo simulations and confidence intervals quantify uncertainty in impact projections for investor disclosure.

2.6 Altruistic Investor System — Algorithm Summary

The Altruistic Investor System integrates the most complex algorithm stack, combining profiling, recommendation, and portfolio optimisation in a single investor-facing experience.

Algorithm	Task	Output
Naive Bayes (NLP)	Information gathering	Investor interest profile from text signals
Random Forest	Areas of influence	Feature-importance ranked SDG affinity map
Fuzzy C-Means	SDG matching	Soft SDG membership scores per investor
Multiple Regression	Project targeting	Predicted engagement probability per project
Hierarchical Clustering / Multivariate Regression	User connections	Peer investor community assignments and co-investment suggestions
Artificial Neural Networks / Collaborative Filtering	User activity	Personalised project feed ranked by predicted engagement
Logistic Regression / Deep Learning	Portfolio interaction	Investment propensity scores and multi-step portfolio behaviour predictions

Glossary of Terms

Term	Definition
DBSCAN	Density-Based Spatial Clustering of Applications with Noise — clusters data points by density, identifying outliers as noise without requiring a pre-set cluster count
DXA	Device-independent units used in OOXML (1440 DXA = 1 inch); the unit system for Word document layout measurements
EM Algorithm	Expectation-Maximisation — iterative statistical optimisation algorithm used to find maximum likelihood estimates in models with latent variables
Fuzzy C-Means	A clustering algorithm allowing data points to belong to multiple clusters with degrees of membership, rather than hard assignment
KNN	K-Nearest Neighbors — classification algorithm that assigns labels based on the majority class among the K closest training examples in feature space
Lasso Regression	Least Absolute Shrinkage and Selection Operator — linear regression with L1 regularisation that performs automatic feature selection by shrinking coefficients to zero
NLP	Natural Language Processing — the branch of AI concerned with enabling computers to understand, interpret, and generate human language
PhREVO Coin	The platform's native digital asset used for impact investment transactions and value transfer between investors, projects, and the Impact Exchange
SDG	Sustainable Development Goals — the 17 global goals established by the United Nations in 2015 as a universal call to action to end poverty, protect the planet, and ensure prosperity for all
SVM	Support Vector Machine — supervised learning algorithm that finds the optimal separating hyperplane between classes in high-dimensional feature spaces
Total Project Value	The composite score produced by the Project Verification System (Score 1 + Score 2 + Score 3) that determines a project's tradeable value on the Impact Exchange
UIC	Universal Impact Certificate — a financial instrument in the PhREVO ecosystem that certifies verified social and ecological impact and can be traded on the Impact Exchange