SUPPLEMENTARY MATERIAL S2

Technical Specifications

For PIU Data Integration, Dashboards, and Forecasting

Purpose: Provide technical guidance for building a Preparedness Intelligence Unit (PIU) information layer that integrates surveillance, laboratory, and climate/environmental feeds into a unified dashboard and forecasting workflow.

Audience: Public Health Emergency Operations Centre (PHEOC) informatics teams, District Health Information Software 2 (DHIS2) administrators, surveillance officers, data scientists, and partners supporting national implementation.

Scope: Reference architecture, minimum technical baseline, data integration protocols, sample Application Programming Interface (API) patterns for DHIS2, and example forecasting workflows in R and Python.

1. Minimum Technical Baseline

1.1 Core Components

- DHIS2 instance with electronic Integrated Disease Surveillance and Response (eIDSR) module (or equivalent event-based surveillance dataset)
- Secure data integration service (Extract, Transform, Load (ETL) / Extract, Load, Transform (ELT)) running on-premise or cloud
- Analytics store (PostgreSQL recommended) and optional time-series store for streaming signals
- Dashboard layer (web-based) with Role-Based Access Control (RBAC)
- Audit logging, backup, and recovery procedures

1.2 Minimum Security Controls

- RBAC mapped to PIU roles and incident roles — see Supplementary Material S3 (Table 2) for the full access matrix
- Least-privilege permissions for service accounts (read-only where possible)
- Transport Layer Security (TLS) for data in transit; encryption at rest for databases and backups
- Central audit logs for data access and dashboard logins
- Data minimization: default to aggregated views; restrict Personally Identifiable Information (PII) to authorized epidemiologists

2. Dashboard Architecture

2.1 Reference Architecture

A PIU dashboard stack follows a three-layer model: (1) Sources, (2) Integration and Analytics, (3) Products and Dissemination.

Layer	What it contains	Examples
Sources	Surveillance, laboratory, climate, CRVS, logistics, mobility signals	eIDSR/DHIS2, Laboratory Information System (LIS), rainfall feeds, stock systems
Integration	ETL, validation, deduplication, identity resolution	Airflow, dbt, custom ETL service

https://doi.org/10.37432/jieph-d-25-00266

Layer	What it contains	Examples
Analytics	Indicators, anomaly detection, forecasting, scoring	Python/R pipelines, SQL models
Products	Dashboards, briefs, alerts, exports	Web dashboards, weekly brief, alert Short Message Service (SMS)/email

2.2 Role-Based Views

- PIU Epidemiologist view: line lists (when authorized), outbreak curves, transmission risk maps, case definitions, and investigation status
- Data Scientist view: data quality panels, model diagnostics, forecast performance (Brier score), feature availability
- Operations Planner view: response readiness, Rapid Response Team (RRT) status, stock levels, facility readiness, dispatch tracking
- Communications view: approved aggregated summaries and key messages (no PII)

2.3 Core Dashboard Modules (Minimum Viable)

Module	Contents
Daily situational awareness	Key alerts, cases, deaths, positivity, geographic clustering
7-1-7 monitoring	Detection, notification, response timelines by priority disease
Data quality	Completeness, timeliness, duplicates, missing geo, lab linkage rates
Forecast panel	14-day and 30-day probabilistic forecasts with uncertainty bands
After-Action Review (AAR) tracker	Issues, owner, due date, status, verification evidence

3. DHIS2 API Integration (Patterns and Examples)

This section shows common query patterns for DHIS2. Adapt endpoints and identifiers to your national configuration.

3.1 Authentication and Service Accounts

- Create a DHIS2 service account for PIU integration with read-only access to required datasets
- Use basic authentication (auth) over Hypertext Transfer Protocol Secure (HTTPS), or Open Authorization 2.0 (OAuth2) if configured
- Rotate credentials and store secrets in a vault or secured environment variable store

3.2 Common Endpoints

Purpose	Endpoint
`Analytics (aggregate)`	`/api/analytics`
`Events (line-list)`	`/api/events`
`Tracked entity instances (if used)`	`/api/trackedEntityInstances`

https://doi.org/10.37432/jieph-d-25-00266

Purpose	Endpoint
Organisation units	/api/organisationUnits
Data elements and indicators	/api/dataElements, /api/indicators

3.3 Example: Pull Aggregated Indicators

Example request (Uniform Resource Locator (URL) pattern):

```
GET
/api/analytics?dimension=dx:IND_1;IND_2&dimension=pe:LAST_12_WEEKS&dimension=ou:LEVEL-2&displayProperty=NAME
```

Example response handling (pseudo-code): Parse headers and rows. Store as a fact table with fields: indicator_id, period, org_unit_id, value, extracted_at.

3.4 Example: Pull Event-Level Signals for Anomaly Detection

Example request (URL pattern):

```
GET /api/events?program=PROGRAM_UID&startDate=2026-01-01&endDate=2026-02-28&orgUnit=OU_UID&includeDeleted=false&paging=false
```

Minimum fields to store: event_id, event_date, org_unit_id, disease_code, case_classification, sex, age_group, lab_status, reported_at.

3.5 Data Mapping Checklist

- Define the national list of priority diseases and map DHIS2 codes to standard labels
- Define event date, report date, and confirmation date fields used for alert-to-response time (ART) timelines
- Define geo hierarchy (district, chiefdom, facility) and the authoritative org unit IDs
- Define linkage keys to laboratory systems (specimen ID, patient ID where lawful)

4. Data Integration Protocols

4.1 Ingestion and Validation

- Extract: schedule pulls (daily) and event pulls (near-real time where possible)
- Validate: schema checks, date ranges, mandatory fields, value ranges
- Standardize: disease codes, age bands, location IDs, case classifications
- Deduplicate: event IDs, repeated alerts, repeated lab results
- Load: write to staging, then curated analytics tables

4.2 Linkage Protocols (Surveillance to Laboratory)

- Link using specimen ID or lab accession number where available
- Record linkage status: linked, pending, not linkable
- Maintain a linkage audit table with timestamps and linkage method

4.3 Data Quality Key Performance Indicators (KPIs) (Minimum)

KPI	Definition
Completeness	% of records with required fields (date, location, disease)
Timeliness	Median days from event date to report date

https://doi.org/10.37432/jieph-d-25-00266

KPI	Definition
Consistency	% of records with valid codes and classifications
Lab linkage	% of suspected cases with laboratory linkage when applicable

5. Forecast Models (R/Python Workflows)

Forecasting supports decision lead-time. The PIU can start with simple models and scale to more advanced approaches as data quality improves.

5.1 Forecast Targets

- Case counts by district for priority diseases
- Threshold breach probability (e.g., probability cases exceed an alert threshold)
- Resource demand proxies (Oral Rehydration Salts (ORS) demand, bed occupancy proxies) where data exist

5.2 Minimum Viable Model Options

Model type	Description
Baseline seasonal model	Moving average + seasonal decomposition; suitable for starting point
Count model	Negative binomial regression with lag terms; accounts for over-dispersion in case counts
State-space model (advanced)	Bayesian hierarchical model for district-level pooling; stronger uncertainty quantification

5.3 Example Workflow (Python)

- Extract curated weekly counts from analytics store
- Split training and evaluation periods (rolling origin evaluation)
- Fit model; generate probabilistic forecasts for 14 and 30 days
- Compute Brier score for threshold events and calibration plots
- Publish forecasts to dashboard with uncertainty bands

5.4 Example Workflow (R)

- Build a time-series object by district and disease
- Fit a baseline seasonal model; compare against a count model
- Generate predictive distribution and evaluate with proper scoring rules
- Document assumptions and data limitations in the weekly brief

6. Operations and Maintenance

- Run daily ETL monitoring with alerts on failures
- Maintain a data dictionary and a change log for indicators
- Quarterly security review and access recertification
- Backup schedule: daily incremental, weekly full, offsite copy
- Disaster recovery test at least twice per year

https://doi.org/10.37432/jieph-d-25-00266

For role-based access controls and data governance requirements that govern this technical infrastructure, see Supplementary Material S3. For performance monitoring metrics that this infrastructure must support (Brier scores, ART, data quality KPIs), see Supplementary Material S6.

List of Abbreviations

Abbreviation	Full Term
AAR	After-Action Review
API	Application Programming Interface
ART	Alert-to-Response Time
Brier score	Mean squared difference between predicted probability and observed outcome (forecast accuracy metric)
CRVS	Civil Registration and Vital Statistics
CSV	Comma-Separated Values
dbt	Data Build Tool (open-source data transformation tool)
DHIS2	District Health Information Software 2
eIDSR	Electronic Integrated Disease Surveillance and Response
ELT	Extract, Load, Transform
ETL	Extract, Transform, Load
HTTPS	Hypertext Transfer Protocol Secure
KPI	Key Performance Indicator
LIS	Laboratory Information System
OAuth2	Open Authorization 2.0 (authentication standard)
ORS	Oral Rehydration Salts
PHEOC	Public Health Emergency Operations Centre
PIU	Preparedness Intelligence Unit
PII	Personally Identifiable Information
Python	Python (programming language)
R	R (statistical programming language)
RBAC	Role-Based Access Control
RRT	Rapid Response Team
SMS	Short Message Service
SQL	Structured Query Language
TLS	Transport Layer Security
URL	Uniform Resource Locator
7-1-7	Detect within 7 days, notify within 1 day, initiate early response within 7 days

https://doi.org/10.37432/jieph-d-25-00266

https://doi.org/10.37432/jieph-d-25-00266