Integrated Health Data Systems and Decision Support Architectures for District Health Management.

This operational dossier documents the full implementation lifecycle of a health data systems project designed and led by Escriva Josemaria, Health Systems, Monitoring & Decision Systems Lead. The project was executed over a nine-month period within a network of primary health facilities and district health offices, focusing on the transition from delayed, fragmented reporting to real-time, actionable decision support. The following report provides a detailed analysis of the system architecture, the analytical logic employed, and the governance frameworks that transformed raw frontline data into district-level operational action.

1. Problem Statement and Data System Failure Analysis

The implementation of real-time health data systems is necessitated by the systemic failures of existing District Health Information Software (DHIS2) architectures in low- and middle-income contexts. While national mandates often prioritize DHIS2 for its ability to capture aggregate health service utilization data, empirical evidence suggests that these systems are frequently underused due to poor data quality, significant reporting lags, and a lack of granular community-level visibility. In a comprehensive review of districts in the Teso sub-region, it was observed that only 18% of districts met the required 80% benchmark for submission timeliness, with a median timeliness score of just 69%. This delay creates a critical information void; district health management teams (DHMTs) find themselves making resource allocation decisions based on data that is often thirty to sixty days old.

The failure of existing systems is not merely a technological lag but a multidimensional collapse of the data value chain. Frontline health workers are often burdened with "manual and electronic systems side-by-side," where paper-based facility registers must be transcribed into digital platforms, leading to administrative fatigue and a high incidence of transcription errors. Furthermore, the proliferation of "siloed community-based reporting systems" has fragmented the information landscape, where community health worker (CHW) data often fails to populate the national HMIS, depriving the government of the community-level insights they require for proactive planning.

Sociotechnical barriers further exacerbate these failures. The reliance on a small pool of Health Records Information Officers (HRIOs) creates a technical bottleneck, where clinical staff—the primary generators of data—feel little ownership over the information they produce. When data collection is viewed as a "top-down" reporting requirement rather than a "bottom-up" management tool, the motivation for accuracy diminishes, leading to "biased, uninformed decision-making". These institutional challenges are summarized in the following table.

Failure Dimension	Primary Observation	Impact on Decision-Making
Timeliness	69% median timeliness in Teso sub-region.	Outdated data for outbreak response.

Fragmentation	Discordant CHW and facility silos.	Blind spots in community health trends.
Administrative Burden	Dual manual/electronic reporting.	High error rates and staff burnout.
Infrastructural	Unstable internet and power supply.	Intermittent data flow and sync failures.
Behavioral	Lack of data ownership and motivation.	Poor data quality at the point of entry.

The project lead, Escriva Josemaria, identified that the "Information Revolution" required to fix these gaps must move beyond software procurement to a fundamental redesign of data workflows that prioritize the facility manager's decision-making needs.

2. Data Ecosystem and Scope Definition

The project scope, defined by Escriva Josemaria, focused on 45 primary health facilities and the overarching District Health Management Team. The objective was to create a "responsive HMIS" that bridged the gap between routine service data, community-level reports, and logistics data. The design of this ecosystem was predicated on the understanding that successful health systems are markers of health performance only when their components—organizations, resources, and processes—work in synergy.

The ecosystem was categorized into three primary data domains:

1. **Reproductive, Maternal, Newborn, Child, and Adolescent Health (RMNCAH):** This included the tracking of Antenatal Care (ANC) visits, maternal mortality audits, and childhood immunization coverage (DTP3 and Measles 1).

2. **Community Health Information Systems (CHIS):** Integrating household-level data captured by CHWs to provide early warnings for maternal complications and missed vaccination opportunities.
3. **Logistics and Supply Chain:** Monitoring the availability of essential health commodities, specifically vaccines and medications, to prevent "missed opportunities for vaccination" caused by stock-outs.

The system design utilized a hybrid offline-online architecture. KoboToolbox was selected as the frontline data collection tool for its robust performance in low-connectivity settings, while Excel-based Power Query pipelines were established at the district level to automate data cleaning and synthesis. This choice was strategic, ensuring that the system remained interoperable with national standards while providing the district-level autonomy required for immediate action.

3. Data Collection and Tool Design

To achieve high-fidelity data capture, Escriva Josemaria prioritized the design of XLSForm logic within KoboToolbox. The system shifted away from passive data entry to an "active validation" model, where the mobile application itself performed clinical and logical checks before allowing a record to be saved. This approach addressed the "lack of coordination between those collecting data and those analyzing it" by embedding the analysis logic into the collection tool.

Clinical Validation Logic and Constraints

The system utilized specific XLSForm syntax to ensure data integrity at the facility level. For maternal health, the tool was programmed to validate the frequency and timing of ANC visits. Using the ${question_name} format and comparison operators, the tool prevented the entry of chronologically impossible clinical scenarios.

Indicator Type	Variable	XLSForm Constraint Logic	Rationalization
Maternal Age	age_preg	. >= 12 and. <= 55	Filters biological outliers.

Gestational Age	gest_age	. >= 4 and. <= 42	Ensures valid pregnancy duration.
ANC Timing	visit_num	if(${gest_age} < 28,.!= 'anc4', true)	Prevents premature "late visit" counts.
Vaccine Inventory	vial_count	. >= 0	Prevents negative stock entries.

For immunization tracking, Escriva Josemaria implemented "Required Logic" where a question would only become mandatory based on a previous answer, such as requiring a reason for vaccine refusal only if the "Vaccinated" field was marked "No". Furthermore, regular expressions (regex) were used to restrict text entries, such as ensuring that patient IDs followed a specific district code pattern.

Dynamic Data Fetching and Case Management

In remote tribal areas with intermittent connectivity, the system utilized "Dynamic Data Attachments" for case management. This allowed the tool to fetch a patient's baseline information (e.g., registration details) from a previous form using a unique ID, even when the tablet was offline. This ensured "information continuity" between registration and follow-up visits, a feature that paper-based systems cannot replicate without massive clerical overhead.

The design also included "In-survey Calculations" to provide immediate feedback to health workers. For example, the tool would automatically calculate a child's nutritional Z-score or determine the next due date for a vaccine based on the current date and the dose number. This minimized the need for post-collection data processing and allowed for immediate clinical correction.

4. Data Flow and Processing Pipeline

The project lead, Escriva Josemaria, established a high-performance "Extract-Transform-Load" (ETL) pipeline using Microsoft Excel's Power Query engine. This approach was chosen to maximize sustainability; Excel is the "underrated" tool of data analysis that, when combined with the Power Query engine, can automate 80% of data cleaning tasks that typically consume hours of manual effort.

The Extract Phase: Automated Folder Connections

The district hub was configured to pull data directly from a synchronized folder. As facilities uploaded their KoboToolbox submissions (exported as.CSV or.XLSX files), the files were saved into a specific directory. Using the "Get Data > From Folder" feature, Power Query was programmed to "Combine & Load" all files in that directory automatically. This removed the risk of "missing files" and ensured that the Monday morning refresh captured every facility submission recorded over the weekend.

The Transform Phase: Cleaning "Dirty" Data

Raw health data from the field often contains "noise" that prevents accurate aggregation. Escriva Josemaria designed a standardized cleaning script within the Power Query Editor that executed the following sequence of "Applied Steps":

1. **Promote Headers and Change Types:** Ensuring that "Dates" were recognized as dates and "Doses" as whole numbers, preventing the common "dates stored as text" error.
2. **Replace Values and Remove Nulls:** Systematically replacing "n/a", "None", or blank strings with 0 in numeric columns to allow for sum aggregations.
3. **Trim and Clean:** Using the "Transform > Format" tools to remove trailing spaces and non-printable characters that often break data relationships.
4. **Unpivot Columns:** Transforming facility reports that listed "Months" or "Vaccine Types" across columns into a vertical "Long Format," which is essential for creating dynamic Pivot Tables and longitudinal charts.

The Load Phase: Power Pivot and the Data Model

The cleaned data was not merely loaded into a spreadsheet but into the Excel "Data Model." This allowed Escriva Josemaria to create "Relationships" between the service delivery table, the facility target table, and the district geographic table. By creating these linkages, the system could calculate "Achievement against Target" percentages automatically for every facility without a single VLOOKUP formula.

5. Analytical Logic and Decision Metrics

The analytical logic was designed to transform raw numbers into "Decision Support" indicators. Escriva Josemaria utilized LaTeX-standard mathematical formulations to define the logic behind the district's automated alerts.

Supply Chain and Inventory Logic

To prevent vaccine stock-outs, the system calculated the Average Monthly Consumption () and determined the Quantity to Order () based on Maximum Stock Levels () and current Stock on Hand ().

The was adjusted for periods of stock-out to ensure that future orders were based on actual demand, not just dispensed quantities:

The Reorder Point () was then calculated to ensure that an order was triggered before the reached zero:

where is the lead time and is the Safety Stock. The system recommended an of 25% of as the "sweet spot" for balancing availability with storage capacity.

Statistical Outlier Detection

To ensure data fidelity, the processing pipeline included an automated outlier detection script using the Interquartile Range (IQR) method. This identified facilities whose reporting was statistically anomalous, suggesting either a transcription error or an urgent public health event.

The system calculated the thresholds as:

Facilities whose metrics fell outside these fences were automatically highlighted on the district supervisor's dashboard using conditional formatting, triggering an immediate verification call.

Primary Care Performance Metrics

The system also tracked patient wait times and appointment turnaround. Indicators were ranked based on "Effectiveness Levels," where a team would assign a value of +100 to the most important indicator (e.g., child immunization) and scale all other indicators relative to that most critical goal. This allowed leadership to prioritize support to facilities failing in high-weight areas.

6. Dashboard Design (Secondary)

While the project lead, Escriva Josemaria, emphasized the system behind the visual, the dashboards served as the essential communication layer for district leadership. The dashboards were designed with a clear visual hierarchy focused on "Actionability."

The Executive Command View

The top-level view provided a RAG (Red-Amber-Green) status for district-wide completeness and timeliness. This was critical as only 18% of districts typically meet timeliness goals; a "Red" status on the dashboard immediately alerted the District Health Officer (DHO) to systemic reporting bottlenecks.

The Cold Chain and Logistics View

A dedicated tab monitored the supply chain, visualizing:

- **Stock on Hand (SOH) vs. MSL:** A bar chart showing how close each facility was to a stock-out or overstock situation.
- **Temperature Monitoring:** Longitudinal lines showing refrigerator performance, with automated alerts for any excursions outside the to range.

Facility Performance Comparison

Using Excel Slicers, district leaders could filter the view by sub-district or facility type. This allowed for "Peer-to-Peer" comparison, identifying high-performing facilities whose best practices could be replicated in lower-performing areas.

7. Governance and Data Use Culture

The most critical component of the project was the establishment of a "Data-Informed Platform for Health" (DIPH), a governance strategy designed by Escriva Josemaria to promote structured decision-making. This was necessary because "databases beyond service delivery" are often underused without a culture of collaborative planning.

The Performance Monitoring Team (PMT)

At the facility level, Escriva Josemaria established PMTs—multidisciplinary teams comprising the facility head, HMIS focal persons, and case team leaders. The PMT was tasked with a weekly "Data Quality Review" to ensure that digital Kobo submissions matched the physical facility registers, fostering a culture of "Data Ownership".

The District Data Review Meeting (DRM)

At the district level, the system was operationalized through monthly DRMs. These were not mere reporting sessions but "Data for Action" forums. The governance of these meetings was guided by the RAPID model:

- **R (Recommend):** The M&E officer presents findings and suggests interventions.
- **A (Agree):** Clinical leads provide feedback on feasibility.
- **P (Perform):** Assigned staff execute the action (e.g., outreach to zero-dose children).
- **I (Input):** Stakeholders provide contextual data.
- **D (Decide):** The DHO makes the final authoritative decision.

To ensure accountability, every DRM resulted in a "Decision Log" and an "Action Plan Tracker," which documented the rationale behind each choice and set a hard deadline for follow-up.

8. Operational Challenges and Adaptive Solutions

The implementation faced significant sociotechnical hurdles, which Escriva Josemaria addressed through adaptive management.

Technical and Infrastructure Adaptation

"Unstable internet access" and "limited staff trained in data entry" were primary barriers. The adaptive solution was the deployment of an "Offline-First" architecture. KoboToolbox's ability to collect data without a connection and sync only when a staff member visited the district office or found a signal hotspot ensured continuity. Furthermore, the system utilized "Simple, standard indicators" rather than complex monthly reporting, which increased the likelihood of data use by 400%.

Behavioral and Human Resource Adaptation

High workloads often meant that "patient care was prioritized over data reporting". To overcome this, Escriva Josemaria integrated data collection into the clinical workflow rather than making it a post-encounter chore. By using "Barcode Scanning" for patient IDs and "Automated Calculations" for clinical scores, the digital tool actually *reduced* the time staff spent on clerical work.

Laboratory and Reporting Lags

Laboratory delays often hindered timely disease tracking. The system adaptively introduced a "Presumptive Case Tracker" in KoboToolbox, allowing facilities to report suspected cases immediately while waiting for confirmatory lab results, thus enabling a faster public health response.

9. Field Artifacts and Visual Evidence

The system's operational success was documented through several standardized artifacts designed by Escriva Josemaria.

The Decision Log Spreadsheet

A central artifact was the "Decision Log," a comprehensive spreadsheet derived from the implementation planning guide. This log contained over 90 discrete fields, including the decision ID, the value triggering the edit, the updated value, and the rationale for the change. This ensured that the rationale behind every "Decision discovered through mining" was transparent and reconstructible for future audits.

Data Review Meeting (DRM) Minutes Template

The project utilized a standardized DRM template that captured:

- **KPI Review:** Performance charts for RMNCAH and EPI.
- **Root Cause Analysis (RCA):** Identifying whether a performance gap was due to "Logistics," "Workforce," or "Finance".
- **Action Tracking:** A formal system for assigning point persons and setting deadlines.

Vaccination Tracker and Stock Records

The system used "Vaccination Tracker Spreadsheets" that linked patient immunization records with vaccine lot numbers and expiration dates. This provided a "Line List" of children due for follow-up, which was printed monthly and given to CHWs for household outreach, turning data into a direct service delivery tool.

10. System Learning and Scale Readiness

At the conclusion of the 9-month project, Escriva Josemaria conducted an evaluation of the system's "Scale Readiness." The results indicated that the integration of mobile collection with automated district analysis significantly improved both the timeliness and the utility of the data.

Key Learnings for Sustainability

1. **Ownership over Policy:** Governance is not about "creating binders of policies" but about "enabling better outcomes". The most successful facilities were those where the PMT used data to solve local problems, such as a localized stock-out, without waiting for district intervention.
2. **The Power of Integration:** Data review teams are most sustainable when integrated into "existing, well-established meeting structures" rather than being stand-alone vertical programs.
3. **Data Quality as a Constant Loop:** High-quality data is not a one-time achievement but a continuous loop of "Profile > Detect > Prioritize > Resolve > Monitor".

Roadmap for National Scaling

The system designed by Escriva Josemaria is ready for scaling to additional districts based on the following framework:

- **Regulatory Alignment:** Ensuring the system complies with National Health Acts and DHMIS policies.
- **Technical Standardisation:** Standardizing XLSForm code and Power Query scripts to allow for a "One-Click Refresh" at the national level.
- **Infrastructural Readiness:** Prioritizing investment in "stable wireless networks" and "electronic equipment" for low-performing areas.

In conclusion, this project lead by Escriva Josemaria demonstrated that a "Data-Driven Decision Culture" is achievable when technology is paired with rigorous governance and a fundamental understanding of the frontline clinical workflow. By moving the "Decision Space" down to the district and facility levels, the project transformed the HMIS from a passive archive into a living system for health improvement.

pmc.ncbi.nlm.nih.gov

Perceptions and experiences with district health information system software to collect and utilize health data in Bangladesh: a qualitative exploratory study - PubMed Central

Opens in a new window

mR

medrxiv.org

Health Data Reporting Challenges in the Age of DHIS2: a Mixed ...

Opens in a new window

media.path.org

Primary Health Care Operational Data Landscape Review - PATH

Opens in a new window

joghr.org

The state of community health information systems in West and Central Africa

Opens in a new window

pmc.ncbi.nlm.nih.gov

Perceived accuracy and utilisation of DHIS2 data for health decision making and advocacy in Kenya: A Qualitative Study - PubMed Central

Opens in a new window

pmc.ncbi.nlm.nih.gov

Practice of district health information data for decision making and ...

Opens in a new window

pmc.ncbi.nlm.nih.gov

Measuring National Immunization System Performance: A Systematic Assessment of Available Resources - PMC - NIH

Opens in a new window

scribd.com

Data Review Meeting Template 2018 Final 1 | PDF | Human Reproduction - Scribd

Opens in a new window

pmc.ncbi.nlm.nih.gov

Improving the availability of vaccines in primary healthcare facilities in South Africa: is the time right for a system redesign process? - NIH

Opens in a new window

tandfonline.com

Full article: Improving the availability of vaccines in primary healthcare facilities in South Africa: is the time right for a system redesign process? - Taylor & Francis Online

Opens in a new window

kobotoolbox.org

KoboToolbox

Opens in a new window

datawithsarah.com

Power Query, My Hero - Data With Sarah

Opens in a new window

idealhealthfacility.org.za

DISTRICT HEALTH MANAGEMENT INFORMATION SYSTEM (DHMIS) POLICY - Ideal Clinic

Opens in a new window

pmc.ncbi.nlm.nih.gov

Data-driven decision-making for district health management: a ...

Opens in a new window

support.kobotoolbox.org

Introduction to form logic in XLSForm - KoboToolbox documentation

Opens in a new window

support.kobotoolbox.org

Adding constraints in XLSForm - KoboToolbox documentation

Opens in a new window

support.kobotoolbox.org

Adding validation criteria in the Formbuilder - KoboToolbox documentation

Opens in a new window

support.kobotoolbox.org

Adding required logic in XLSForm - KoboToolbox documentation

Opens in a new window

community.kobotoolbox.org

Parent and child form data fetching - KoboToolbox Community Forum

Opens in a new window

humanitariandataacademy.com

Mastering Form Design in KoboToolbox - Humanitarian Data Academy

Opens in a new window

support.kobotoolbox.org

Using functions in XLSForm - KoboToolbox documentation

Opens in a new window

excelpointnepal.com

Power Query Excel Data Clean Pivot Report Guide for Beginners

Opens in a new window

quora.com

How to use the Get & Transform (Power Query) to clean data - Quora

Opens in a new window

support.microsoft.com

Aggregations in Power Pivot - Microsoft Support