Integrated Data Warehouse

A central data warehouse unifies client records across disparate systems, putting comprehensive data into a usable format for cohort and predictive analytics

THE OPPORTUNITY

A considerable data challenge facing the HIV response is a lack of countrywide client-line longitudinal data. Such data enables us to calculate month retention, suppression, and mortality rates over different time intervals, and conduct effective data quality interventions, such as *anomaly detection*. Specifically, longitudinal data analysis can flag *anomalies*—values not explained by trend and seasonality. Longitudinal client-line data also support *granular analyses* through geospatial layering or clustering by client characteristics (sex, age) or subpopulations, such as key populations (KPs) and adolescent girls and young women (AGYW).

Data warehousing of client-line data allows for multivariate analyses of time-series data and routinizes analyses for high-frequency reporting, HIV case-based surveillance, and HIV outcomes measurement, such as cohort and survival analyses. Data warehousing also enables improved deduplication of client records. It is foundational in the application of artificial intelligence and machine learning of HIV service data, through which the efficacy and impact of differentiated care models may be better assessed.

THE DATA.FI SOLUTION

Data.FI works with U.S. Agency for International Development (USAID) Missions and host country governments to build and enhance data warehouses (DWHs) that align to user needs and maximize the usability of all available data for decision making. By enabling data exchange through interoperability, data will be transformed for input into analytical systems to enable case-based surveillance and data visualization. This solution includes the following:

Architecting, networking, interoperating: Data.FI aligns stakeholders' visions for the DWH, building consensus on and documenting the data standards and governance structure needed for centralizing health information. We outline data security protocols that align with country regulations, including how data are transmitted and stored, and who has access. The DWH pulls data from multiple source systems, including client-line, facility-line, and aggregated data. Using an Application Programming Interface (API), we then use extract, transform, and load (ETL) routines to populate the appropriate data marts within the DWH with source system data. In building the interoperability layer, we use internationally recognized standards (e.g., OpenHIE) for health data exchange in resource-constrained environments.

Deduplicating client records: We match and de-duplicate client records within the electronic medical record (EMR) and across source systems, creating a unified and longitudinal client record that includes information on testing, treatment, medication dispensing, and viral suppression. This consolidation enables cohort analyses that unlock accurate retention data, improve case-based surveillance, and define progress toward epidemic control.

Automating program analytics and customizing reports: We automate descriptive and predictive analytics, and customize data visualizations and reports to end users to support continuous quality improvement. The availability of comprehensive client-line data allows us to predict client retention and adherence challenges before they occur so they can be prevented.

Training local staff to maintain the system: Our solution underscores sustainability; we work with USAID and governments to identify long-term hosting and maintenance strategies from the beginning, and partner with and train local technology partners and government personnel on system maintenance and use.
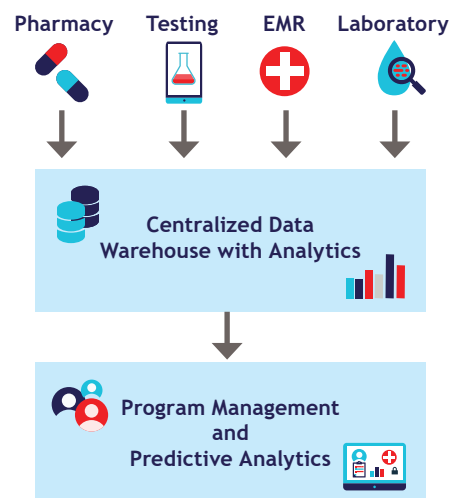
WHAT IS THE IMPACT?

In Kenya, through CDC/KeHMIS II, Palladium developed a national DWH for the Ministry of Health for use in annual cohort reporting and granular high-frequency reporting. The number of EMR facilities that uploaded data to the DWH at least once increased from 340 facilities in 2017 to more than 1,000 facilities in 2019, representing a 180 percent growth in data uploads over that period. This growth has led to a massive increase in the volume of data in the DWH. Within six months, the number of records increased by 400,000, from 1.4 million to 1.8 million records. This increase has improved the representativeness and value of data for decision making by the national HIV program.

In South Africa, Zambia, and Zimbabwe, Data.FI partner Right to Care and Qode's client-centric Knowledge Center integrate and warehouse data from multiple sources (e.g., custom reporting templates, EMRs, lab and pharmacy records). The Knowledge Center is linked to an analytical application to compare data across time and locations. Analytics made possible through the Knowledge Center have improved cascade outcomes. For instance, in Zambia, yields from HIV testing increased from 2 to 10 percent, along with improved data use over the course of one year (weekly yields further increased to 30 percent with additional GIS analytics). In South Africa, routine review of data through the Knowledge Center led to a 6 percent increase in TX_CURR, a 35 percent increase in the average daily case finding rate, and a 32 percent increase in Net New/New ratio between March and May 2019 for 21 priority facilities. This approach was recently evaluated by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) office in South Africa and determined to be a best practice for scale-up.

PUTTING THE SOLUTION INTO ACTION

Data.FI develops solutions with strategic engagement and buy-in from the government, USAID, and other key stakeholders. Data.FI will support countries in doing the following:

- Develop improved health information systems (HIS) architecture and governance structures to ensure data standards, sustainability, and security
- Align stakeholders around data standards, access, and use across sub-system "owners"
- Identify a government-owned hosting solution
- Develop the DWH with linkages to multiple source systems as a comprehensive, secure central repository of client-level data
- Identify and implement an appropriate technological solution to de-duplicate client records
- Conduct analyses and develop visualizations, including dashboards and reports, through a user-centered design process
- Train users and provide capacity strengthening to government and local technology partners to maintain the DWH
- Test and improve the solutions through frequent user feedback

March 2020

FOR MORE INFORMATION

Emily Harris, Data.FI AOR
emharris@usaid.gov
+1-571-309-1266

Jenifer Chapman, Project Director
jenifer.chapman@thepalladiumgroup.com
+1-202-775-9680