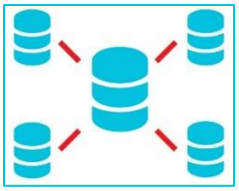


Integrated Data Warehouses

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

THE OPPORTUNITY

Most countries do not have HIV client-line longitudinal data. Such data enables the calculation of retention, viral suppression, and mortality rates over different time intervals, and supports 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). Longitudinal client-line data also promote data quality. Anomalies—values not explained by trend and seasonality—are easily identified.

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. Data warehousing 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.

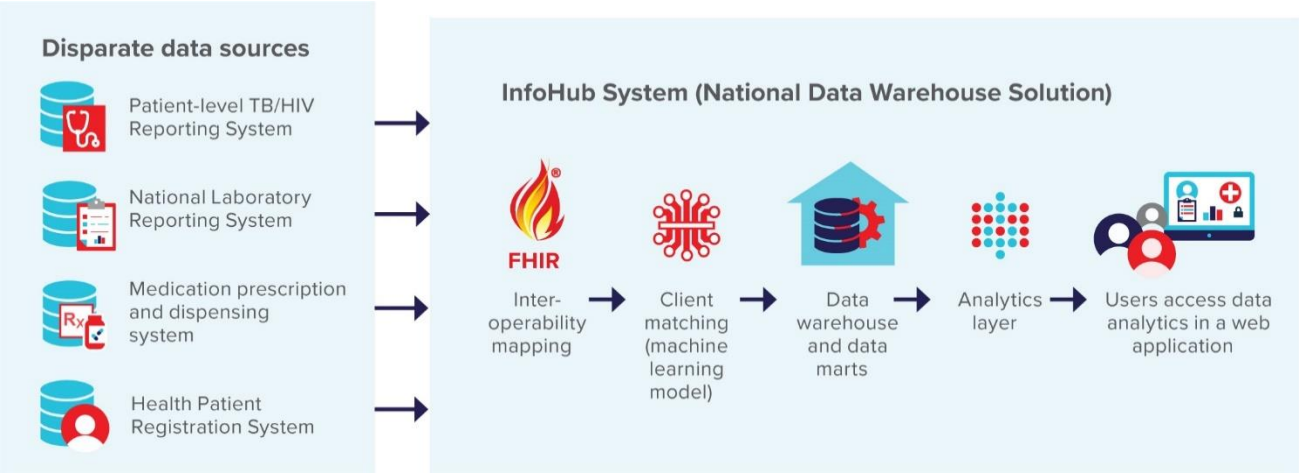
THE DATA.FI SOLUTION

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:

Architecture, 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 deduplicate 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.[1]

[1] Data.FI. (2020). Guidance for Deduplicating Client Records. Washington, DC, USA: Data.FI, Palladium



In South Africa, Data.FI is supporting the National Department of Health to implement a national HIV data warehouse solution called the InfoHub, shown above.

Automating program analytics and customizing reports: We automate descriptive and predictive analytics, and customize data visualization 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 systems: Our solutions underscore 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 South Africa, Data.FI is supporting the National Department of Health (NDOH) to develop a national DWH, as a component of the InfoHub solution. The InfoHub is an information architecture solution for interoperability with the goal of improving data availability for decision making. The InfoHub DWH brings together multiple datasets from disparate sub-systems into an analytical web application that allows a variety of stakeholders to access reports and analyses tailored to their needs. These include treatment outcomes such as cohort reports, viral load monitoring, management and treatment outcomes of HIV-associated co-morbidities, and predictive analytics and anomaly detection. Interactive analytical dashboards will be accessible through a tiered user-access controlled web platform supported by the InfoHub to as many as five thousand NDOH employees and stakeholders countrywide. Access to these curated analyses, tailored for different cadres and geographic units, will enable evidence-informed programming, improved service delivery provision, and adaptive supportive supervision.

A key component to the reliability of the InfoHub analyses is the deployment of machine learning-based client-matching model for records. The InfoHub client-matching model currently matches over 20 million records within the national tuberculosis(TB)/HIV information system and establishes the mechanism for other client records ingested within the InfoHub to be matched and unified. The deduplication report created as an output of the model is shared with facilities to identify duplicate client records for correction

with the central TB/HIV system. This results in more complete and accurate client records, and creates potential for augmenting client records with additional data from the datasets from other subsystems.

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 system (HIS) architecture and governance structures to ensure that data can be optimally accessed and used by all stakeholders for improved program performance
- Ensure HIS architectures and systems maintain industry standards for data exchange, interoperability, privacy, 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 deduplicate 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, with an emphasis on long-term sustainability
- Test and improve the solutions through frequent user feedback

SB-20-07

Data for Implementation (Data.FI) is a five-year cooperative agreement funded by the U.S. President's Emergency Plan for AIDS Relief through the U.S. Agency for International Development under Agreement No. 7200AA19CA0004, beginning April 15, 2019. It is implemented by Palladium, in partnership with JSI Research & Training Institute (JSI), Johns Hopkins University (JHU) Department of Epidemiology, Right to Care (RTC), Cooper/Smith, IMC Worldwide, Jembi Health Systems, and Macro-Eyes, and supported by expert local resource partners.

January 2021

FOR MORE INFORMATION

Contact Data.FI:

Emily Harris, Data.FI AOR
emharris@usaid.gov

Jenifer Chapman, Data.FI Project Director
datafiproject@thepalladiumgroup.com

https://datafi.thepalladiumgroup.com/