World Health Data Hub (WHDH)
case study

The WHO Global Clinical Platform collects clinical data to deepen our understanding of infectious diseases and clinical management during emergencies.
The World Health Organization (WHO) Global Clinical Platform is a secure database comprising of individual-level, anonymized clinical data of hospitalized patients from health facilities worldwide, collected over the period from hospital admission to discharge, death, or transfer. It bridges the gap between epidemiology and clinical care, supplying Member States with critical data and enhancing our understanding of the clinical characteristics among patients with suspected, probable, or confirmed emerging infectious diseases.

Today, the Global Clinical Platform contains data from over 1 million patients, 1000 variables, and 1000 health facilities in over 65 countries. The Platform is rapidly expanding into a Version 2.0, which necessitates substantially scaling the data infrastructure in several areas:

- **Storage Capacity:** Health facilities share data directly to a web-based Electronic Data Capture System or a commercial file-hosting service. The harmonization, curation, and pooling of large volumes of data from diverse sources requires substantial storage capacity.

- **Security:** Contributed data are governed according to the Global Clinical Platform Terms of Use, with rigorous standards for data security and privacy. Solutions must align with prevailing health data storage and sharing policies.

- **Computing Power:** Data processing and analysis is computationally intensive, and standard WHO laptops struggle to handle this load. Scaled computing power makes data processing faster and more frequent, which is crucial in emergencies.

- **Accessibility:** Data pipelines must enable rapid access to data on the Platform via smartphones or tablets to produce dashboards and reports for immediate field use during outbreaks.
Solution

Introducing the World Health Data Hub (WHDH) Data Lake

**Secure, Unlimited Storage:** The WHDH Data Lake is the optimal storage solution for the Global Clinical Platform, securely storing an unlimited volume of data. Data can be securely processed, manipulated, and analysed centrally within the Data Lake environment itself, eliminating the previous workflow to download data to local devices for processing. This enables migration off commercial file-hosting services, which do not meet storage capacity and security requirements.

**Data Lifecycle Management Processes:** The WHDH Data Lake is ideal for WHO teams managing complex data processes across diverse sources. The standard Data Lake contains multiple data quality stages, with the Bronze layer storing raw data, and increasing in data processing level through Silver and Gold. During onboarding, the WHDH team helped clarify processes and configure the Data Lake for the Platform’s specific needs – for example, adding an Iron layer between Bronze and Silver. The Global Clinical Platform team introduced a data quality stage-based communication approach using Data Lake layers, enabling efficient collaboration.

**Scaled Computational Capacity:** The improved computing power on WHDH unlocks speed and scale for data analysis. For example, pooling datasets for the Global Clinical Platform previously took weeks due to limited computing resources and iterative data verification processes. Consequently, data was collated every few months, even though data was collected continuously from health facilities. With WHDH, the Platform can now pool data within days, making quality data available more frequently. Moreover, access to a variety of powerful cloud computers enables computing-intensive exploratory and sensitivity analyses.
Results

A strong foundation for the Global Clinical Platform 2.0

The Global Clinical Platform leverages WHDH services to securely acquire, store, and unlock analytical applications for data. Within three months of onboarding to WHDH, the Global Clinical Platform team reports several early benefits:

- **Seamless Data Migration**: WHDH-customized pipelines efficiently migrated datasets from the previous commercial storage solutions to the Data Lake. Within three months of migration, the Global Clinical Platform team was able to stop actively using commercial storage services, immediately resulting in sufficient and secure data storage.

- **Data Lifecycle Management**: Introducing a clearly defined data lifecycle enhanced quality stage-based data organization and provided a reliable storage system.

- **Enhanced Computing Power**: Significant and immediate improvements to the Global Clinical Platform’s computing power resulted in faster data processing and analytic outputs.

Looking forward, the scaled and flexible data infrastructure on WHDH creates a strong foundation for the Global Clinical Platform to evolve. Future enhancements include leveraging Shiny Server to showcase design dashboards, using xMart to facilitate data sharing across WHO, centralizing a R code repository for interdepartmental collaboration, and expanding pipelines for greater data extraction and contribution.
About the WHO Global Clinical Platform

The Global Clinical Platform is a secure database comprising of individual-level, anonymized clinical data of patients from health facilities across the globe.

The Global Clinical Platform launched in May 2020 at the beginning of the COVID-19 pandemic as a collaboration between the WHE Programme, SRH Department / HRP Programme, and HHS Department, under the guidance of a WHO technical advisory group. Over time, collaboration among WHO departments have expanded to use data collected on the platform to address research questions for infectious diseases such as COVID-19, mpox, viral haemorrhagic fevers and cholera.

Learn more at https://www.who.int/tools/global-clinical-platform

About WHDH

The World Health Data Hub (WHDH) is the end-to-end data suite for WHO. From data collection to use, WHDH is guided by WHO’s data principles to support WHO teams to leverage best-in-class technology addressing data needs:

- Secure and unlimited data storage.
- Member State engagement.
- Collaborative research environment for analytics.
- Public data catalogue.
- Modern data pages and charts.
- Equitable access and internationalization.

Learn more at https://data.who.int/about or email WHDH@who.int

© World Health Organization 2024. Some rights reserved. This work is available under the CC BY-NC-SA 3.0 IGO licence.


Image source. Nurse Ramatu prepares COVID-19 vaccine at a mobile vaccination clinic in Rofunta, Sierra Leone on 6 December 2022 / © WHO / Michael Duff